09/704362 Attachment to Paper 6

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/35316 A2

(51) International Patent Classification[7]: G06F 19/00

(21) International Application Number: PCT/US00/30863

(22) International Filing Date:
10 November 2000 (10.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 09/438,566 | 10 November 1999 (10.11.1999) | US |
| 09/704,362 | 1 November 2000 (01.11.2000) | US |

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

| | |
|---|---|
| US | 09/438,566 (CIP) |
| Filed on | 10 November 1999 (10.11.1999) |
| US | 09/704,362 (CIP) |
| Filed on | 1 November 2000 (01.11.2000) |

(71) Applicants *(for all designated States except US)*: **STRUCTURAL BIOINFORMATICS, INC.** [US/US]; 10929 Technology Place, San Diego, CA 92127 (US). **QUEST DIAGNOSTICS INVESTMENTS INCORPORATED** [US/US]; 300 Delaware Avenue, Wilmington, DE 19899 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RAMNARAYAN, Kalyanaraman** [IN/US]; 11674 Springside Road, San Diego, CA 92128 (US). **MAGGIO, Edward, T.** [US/US]; 18775 Bernardo Trails Drive, San Diego, CA 92128 (US). **HESS, P., Patrick** [US/US]; 54 Via Zeragoza, San Clemente, CA 92672 (US).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman White & McAuliffe LLP, 4250 Executive Square, Suite 700, La Jolla, CA 92037 (US).

(81) Designated States *(national)*: AE, AG, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *Without international search report and to be republished upon receipt of that report.*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/35316 A2

(54) Title: USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS

(57) Abstract: Provided herein are computer-based methods for generating and using three-dimensional (3-D) structural models of target molecules and databases containing the models. The targets can be protein structural variants derived from genes containing polymorphisms. The models are generated using molecular modeling techniques and are used in structure-based drug design studies for identifying drugs that bind to particular structural variants in structure-based drug design studies, for designing allele-specific drugs and population-specific drugs and for predicting clinical responses in patients. Computer-based methods for predicting drug resistance or sensitivity via computational phenotyping are also provided. Databases containing protein structural variant models are also provided.

# USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS

## RELATED APPLICATIONS

Benefit of priority to the following applications is claimed herein: U.S. application Serial No. 09/438,566 to Kalyanaraman Ramnarayan, Edward T. Maggio and P. Patrick Hess, filed November 10, 1999 entitled "USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS FOR DRUG DESIGN AND CLINICAL APPLICATIONS"; and U.S. application Serial No. (Attorney Dkt. No. 24737-1906B) to Kalyanaraman Ramnarayan, Edward T. Maggio and P. Patrick Hess, filed November 1, 2000, entitled "USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS FOR DRUG DESIGN AND CLINICAL APPLICATIONS."

Where permitted the above-noted applications are incorporated by reference in their entirety. Also incorporated by reference in its entiretly is U.S. application Serial No. (attorney docket no. 24737-1906C), filed November 10, 2000, to entitled "USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS."

**Incorporation by reference of Tables provided on Compact Disks**

For US purposes and where permitted, an electronic version on compact disk (CD) ROM of Tables 4 and 5, which set forth coordinates for three-dimensional structures of proteins in the database described herein is filed herewith, and, where permitted and for US purposes, the contents thereof is incorporated by reference in its entirety. Table 4 is the HIV reverse transcriptase coordinates, and Table 5 is the HIV protease coordinates. The files that contain Table 4 are entitled 1906TAB.PC1 and 1906TAB.PC2, created on November 10, 2000, and are 59,538 kilobytes

and 304 kilobytes, respectively. and the file that contains Table 5 is entitled 1906TAB.PC3, created on November 10, 2000, and contains 11,413 kilobytes.

## FIELD OF THE INVENTION

The present invention is related to computer-based methods and relational databases that use three-dimensional (3-D) protein structural models derived from genetic polymorphisms in the areas of computer-assisted drug design and the prediction of clinical responses in patients.

## BACKGROUND OF THE INVENTION

Recent advances in molecular biology, such as the discovery and identification of large numbers of genes and the sequences thereof encoded in the genomes of humans, other mammals and infectious disease agents, have contributed to the identification of a large number of proteins, biological receptors and other macromolecules and complexes that are promising therapeutic targets. Based on the information derived from the gene sequences, the three-dimensional (3-D) molecular structures of the corresponding target proteins or receptors can be determined.

Since 3-D protein structure is related to biological function, structure-based drug design is an increasingly useful methodology that has made a great impact in the design of biologically active lead compounds. Drug designers can design and screen potential new drugs via computational methods, such as docking or binding studies, before actually beginning patient testing. These experiments can be performed *in silico* at a tiny fraction of the clinical cost.

The resulting molecules, while serving as lead compounds, often have unpredictable effects when employed in clinical trials. In addition, it has been observed that existing drugs with known clinical efficacy far often fail to achieve beneficial results when given to particular patients, or

particular subpopulations, such as ethnic groups, of patients. Genetic stratification of a population can be the difference between drug failure and drug approval. Hence there is a need to develop methods to improve the drug discovery process. Therefore, it is an object herein to provide, among a variety of benefits, methods and products that address and solve these problems. In particular, it is an object herein to provide computationally-based methods for drug design, clinical testing protocols, identification of new drug candidates and drug therapies; for predicting drug sensitivity and resistance and other methods.

## SUMMARY OF THE INVENTION

Provided herein are computer-based methods for generating and using three-dimensional (3-D) structural models of target biomolecules, particularly polymorphic and allelic variants. Also provided herein are databases that contain the sequences of such variants and also the 3-D structure of the variants for use with the methods.

Genetic polymorphisms arise, for example, as a result of gene sequence differences or as a result of post-translational modifications, including glycosylation. Hence genetic polymorphisms are manifested as gene products and proteins having variant structures. The variant structures result in differences in biological responses among the originating organisms. These differences in response, include, but are not limited to, differences among patient responses to a particular drug, effective dosage differences, and side effects. With respect to infectious organisms, some polymorphisms may arise that convey resistance or susceptibility to particular drug therapies by the altering the drug target structure.

Structural changes that arise as a result of genetic polymorphisms are not of unlimited variety, since 3-D structure impacts upon function. A knowledge of the repertoire of the fine differences among generally similar

3-D structures of particular proteins will permit design of drugs that bind to the most polymorphisms, drugs that induce the fewest side-effects, and drugs that are more effective against infectious agents. Knowledge of these structures ultimately will permit patient-specific or subpopulation-specific, such as ethic, age, or gender groups, design or selection of drugs.

The methods that are provided are for determining and using 3-dimensional (3-D) protein structures that are derived from genetic polymorphisms to understand differences in biological activity that result from the polymorphisms, and to use this understanding to aid in the identification of potential new drug candidates and drug therapies. Also provided are methods for analyzing 3-D structures of protein structural variant targets derived from genetic polymorphisms to identify common structural features among the variants; methods for identifying structural changes in target proteins that are associated with multiple mutations arising from genetic polymorphisms and correlating this information with biological activity; methods for using clinical data in conjunction with structural variants derived from genetic polymorphisms to understand and predict the pharmacological effects and clinical outcomes for drugs or potential drugs. Also provided are methods for generating 3-D protein structures derived from a given genotype to analyze protein-drug binding *in silico* to predict drug sensitivity or resistance. Also provided are databases that are used in methods provided herein and methods for generating the databases.

In particular, target biomolecules are protein structural variants encoded by genes containing genetic variations, or polymorphisms. 3-D models of the structures of proteins are determined. The models are generated using molecular modeling techniques, such as homology modeling. The resulting models are then used in the methods provided

herein, which include structure-based drug design studies to design and identify drugs that bind to particular structural variants; structure-based drug design studies and to predict clinical responses in patients; and to design drugs that bind to all or a substantial portion of allelic variants of a target, to thereby increase the population of patients for whom a particular drug will be effective and/or to decrease the undesirable side-effects in a larger population.

Hence, computer-based methods of drug design based on target protein structural models derived from genetic polymorphisms are provided. The methods involve obtaining one, preferably two or more amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms, where sequences represent different genetic polymorphisms, and generating 3-D protein structural variant models from the sequences. Structure-based drug design techniques are used to design potential new drug candidates or to suggest modifications to existing drugs based on predicted intermolecular interactions of the drugs or drug candidates with the models. Alternatively, drug molecules can be computationally docked with 3-D protein structural variant models based upon the sequences and energetically refined before performing structure-based drug design studies.

In preferred embodiments, binding interactions between a drug or potential new drug candidate molecules and the structural variants are calculated in order to optimize intermolecular interactions between drug or potential drug molecules and the structural variant models or to select drug therapies for patients by determining a drug or drugs that have favorable binding interactions with the structural variant models.

In other embodiments, the binding interactions are determined by calculating the free energy of binding between the protein structural variant model and a docked molecule; and decomposing the total free

energy of binding based on the interacting residues in the protein active site.

After the protein structural variant models are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models. The conserved structural features can serve as scaffolds or pharmacophore models into which potential drugs or modified drugs are docked. For example, the selected model structures may represent the structural variants resulting from the most commonly occurring genetic polymorphisms or from genetic polymorphisms found in a specific patient subpopulation, such as a particular age group, ethnic or racial group, sex, or other subpopulation. Alternatively, the models may be selected based on clinical information, for example, the structural variants may be derived based on patients receiving a specific treatment regimen or exhibiting a particular clinical response to a given drug or on the duration of a particular drug treatment.

The methods provided herein can be used for predicting clinical responses in patients based on genetic polymorphisms. For example, a structural variant model derived from a subject, such as a human patient, exhibiting a particular genetic polymorphism is generated and screened against a number of reference protein structural variant models derived from genetic polymorphisms of the same gene in other such subjects. In certain embodiments, the reference structures are stored in a database, preferably with observed clinical data associated with the structures, or polymorphisms. The structural variant model from the subject is compared to a reference structures, for example, by database searching, in order to identify reference structural variants that are similar to the model structure derived from the subject. Based on the premise that structurally similar targets will have similar clinical responses, a clinical outcome can be predicted for the patient based on the structures

identified through structural comparison or database searching. This information can also be used in the design and analysis of clinical trials; it can also be used for selecting appropriate therapies for a subject in instances in which the subject is a patient and the protein is a drug target.

The methods are also used to design therapeutic agents that are active against biological targets that have become drug resistant, particularly due to genetic mutations. In certain embodiments, 3-D protein structural variant models are generated for a target protein in which genetic mutations have occurred and against which a given drug is no longer biologically active. The models are compared to 3-D protein structural variant models of the target protein against which the drug has biological activity in order to identify structural differences between the susceptible and resistant targets. The differences can be used to understand the structural contributions to drug resistance, and this information can be utilized in structure-based drug design calculations to identify new drugs or modifications to the existing drug that circumvent the resistance problem.

A computer-based method for identifying compensatory mutations in a target protein is also provided. The method involves obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, where the structure of a form of the target protein that responds to a particular drug, including the active site, has been structurally characterized; generating a 3-D structural model of the mutated protein; comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations; comparing the biological activities of the drug against the mutated protein and the form of the protein that responds to the drug to determine the

effects of the mutations on drug response; and identifying the mutations in the protein that affect biological activity based on the comparisons. The target biolmolecules can also be used in a method referred to herein as computational phenotyping to predict drug sensitivity or resistance for a given genotype. These computer-based method for identifying phenotypes *in silico* are provided. The methods involve obtaining from a patient/specimen, such as a body fluid or tissue sample, including blood, cerebral spinal fluid, urine, saliva, sweat and tissue samples, the amino acid sequence of a target protein; generating a 3-D structural model of the target protein; performing protein-drug binding analyses; and predicting drug sensitivity or resistance based on the protein-drug binding analyses.

Molecular structure databases containing protein structural variant models produced by the methods are also provided. The databases may also contain biological or clinical data associated with the structural variants. The databases can be interfaced to a molecular graphics package for visualization and analysis of the 3-D molecular structural models. In particular, databases containing the 3-D structures of polymorphic variants of selected target genes, particularly pharmaceutically significant genes with pharmaceutically significant gene products, such as proteases and polymerases, including reverse transcriptases, and receptors, such as cell surface receptors, are provided. The databases may be stored an provided on any suitable medium, including, but are not limited to, floppy disks, hard drives, CD-ROMS and DVDs.

Also provided are relational databases for managing and using information relating to genetic polymorphisms. The databases contain 3-D molecular coordinates for structural variants derived from genetic polymorphism, a molecular graphics interface for 3-D molecular structure

visualization, computer functionality for protein sequence and structural analyses and database searching tools. The databases may further include observed clinical data associated with the genetic polymorphism. The databases provide a means to design the allele-specific drugs and also to identify among alleles common or conserved structural features that can serve as the target for drug design.

The databases can also be used for identiication of invariant residues and regions of a target biomoleucle, such as an HIV protease or reverse transcriptase. The identified invariant regions are then used to computationally screen compounds, preferably small molecules by assessing binding interactions. The compounds so-identified serve as candidates for drugs that will be effective for a larger proporation of a population or against a broader range of variants of a pathogen, where the target protein is from a pathogens.

Systems, including computers, containing the databases also are provided herein. Any computer known to those of skill in the art for maintaining such databases is contemplated. User interfaces for accessing and manipulating the databases and content thereof are also provided.

## BRIEF DESCRIPTION OF THE DRAWINGS

**FIG. 1** illustrates a method for creating a protein structural variant relational database.

**FIG. 2** is a flow chart that describes one method used to generate structural variant models derived from genetic polymorphisms and to use the models in structure-based drug design studies.

**FIG. 3** is a flow chart that describes an alternative method used to generate structural variant models derived from genetic polymorphisms and to use the models in structure-based drug design studies.

**FIG. 4** shows the correlation between experimental and calculated changes of binding energy upon ligand modifications in the binding site of NS3.

**FIG. 5** shows a comparison of calculated *versus* experimental binding free energy changes for complexes of the tumor necrosis factor (TNF) receptor with different inhibitors.

**FIG. 6** shows the HIV PR inhibitors approved by the FDA.

**FIG. 7** shows the frequency versus amino acid residue plot of HIV PR.

**FIG. 8** shows frequency analysis of 10591 HIV PR Sequences, where ResNum is the residue number; TotOcc is the total occurrence of the mutation; Dist is the distance of the mutating residue from approximate center of active site (Asp28); WtAA is the amino acid in the wild type protein; NumMut is the number of mutations; and MutList is a list of amino acid mutations.

**FIG. 9** is a block diagram of an exemplary computer.

**FIG. 10** is a graphical representation of a relational database.

**FIG. 11** is a tabulation of the 3-D coordinates of a representative entry in a database that includes 3-D structures.

## DETAILED DESCRIPTION OF THE INVENTION

- **A. Definitions**
- **B. Computer-based methods of drug design based on genetic polymorphisms**
  1. **Methods for obtaining amino acid sequences of a target protein**
  2. **Generation of 3-D protein structural variant models**
     - **a. Homology Modeling**
     - **b. Ab initio generation of 3-D structures**
     - **c. Crystal structures**
  3. **Use of 3-D structural variant models in drug design**
     - **a. Selection of relevant structural variants**
     - **b. Drug design**
     - **c. Computational docking**

d. Free energy of binding studies

C. Applications of computer-based methods

1. Genetic polymorphisms and structure-based drug design
2. Drug resistance
3. Identification of conserved structural features or pharmacophores
4. Identification of compensatory structural changes
5. Clinical Applications

D. Creation of 3-D Structural Polymorphism Databases

1. Exemplary Databases and generation thereof
2. Computer systems and Database

E. Computational phenotyping

## A. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents, patent applications, published patent applications and publications referred to herein are, unless noted otherwise, incorporated by reference in their entirety. In the event a definition in this section is not consistent with definitions elsewhere, the definition set forth in this section will control.

As used herein, polymorphism refers to a variation in the sequence of a gene in the genome amongst a population, such as allelic variations and other variations that arise or are observed. Genetic polymorphisms refers to the variant forms of gene sequences that can arise as a result of nucleotide base pair differences, alternative mRNA splicing or post-translational modifications, including, for example, glycosylation. Thus, a polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. These differences can occur in coding and non-coding portions of the genome, and can be manifested or detected as differences in nucleic acid

sequences, gene expression, including, for example transcription, processing, translation, transport, protein processing, trafficking, DNA synthesis, expressed proteins, other gene products or products of biochemical pathways or in post-translational modifications and any other differences manifested among members of a population. A single nucleotide polymorphism (SNP) refers to a polymorphism that arises as the result of a single base change, such as an insertion, deletion or change in a base.

A polymorphic marker or site is the locus at which divergence occurs. Such site may be as small as one base pair (an SNP). Polymorphic markers include, but are not limited to, restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and other repeating patterns, simple sequence repeats and insertional elements, such as Alu. Polymorphic forms also are manifested as different mendelian alleles for a gene. Polymorphisms may be observed by differences in proteins, protein modifications, RNA expression modification, DNA and RNA methylation, regulatory factors that alter gene expression and DNA replication, and any other manifestation of alterations in genomic nucleic acid or organelle nucleic acids.

As used herein, structural variants proteins refer the variety of 3-D molecular structures or models thereof that result from the polymorphisms. These variants typically arise from transcription and translation of genes containing genetic polymorphisms, but also include diffentially glyocsylated or otherwise post-translationally modified variants that potentially exhibit differential interactions with drugs and drug candidates.

As used herein, binding interactions refer to atomic or physical interactions between molecules including, but not limited to binding free energy, hydrophobic interactions, electrostatic interactions, steric interactions and other interactions that are commonly considered by those of skill in the art to determine the affinity of one molecule to bind to another. Favorable binding interactions refer to binding interactions that promote physical or chemical associations between molecules.

As used herein, a target protein is defined as a protein that is a receptor with which drugs or other ligands, such as small molecule or peptide agonists or antagonists or other proteins or biomacromolecules, such as DNA or RNA, interact to bring about a biological response.

As used herein, structure-based drug design refers to computer-based methods in which 3-D coordinates for molecular structures are used to identify potential drugs that can interact with a biological receptor. Examples of such methods include, but are not limited to, searching of small molecule libraries or databases, conformational searching of a ligand within an active site of identify biologically active conformations or computational docking methods.

As used herein, pharmacogenomics refers to study of the variablity of patient responses to drugs due to inherent genetic differences.

As used herein, computational docking refers to techniques wherein molecules, for example, a ligand and receptor or active site, are fitted together based on complementary interactions, for example, steric, hydrophobic or electrostatic interactions.

As used herein, energetic refinement refers to the use of molecular mechanics simulation techniques, such as energy minimization or molecular dynamics, or other techniques, such as quantum-based approaches, to "adjust" the coordinates of a molecular structural model to bring it into a stable, low energy, conformation. In molecular mechanics

simulations, the potential energy of a molecular system is represented as a function of its atomic coordinates along with a set of atomic parameters, called a forcefield. Energy minimization refers to a method wherein the coordinates of a molecular conformation are adjusted according to a target function to result in a lower energy conformation. Molecular dynamics refers to methods for simulating molecular motion by inputting kinetic energy into the molecular system corresponding to a specified temperature, and integrating the classical equations of motion for the molecular system. During a molecular dynamics simulation, a system undergoes conformational changes so that different parts of its accessible phase space are explored.

As used herein, clinical data refers to information obtained from patients pertaining to pharmacological responses of the patient to a given drug, including, but not limited to efficacy data, side effects, resistance or susceptibility to drug therapy, pharmacokinetics or clinical trial results.

As used herein, patient histories, include medical histories and other any information, such as parental medical histories, dates and places of birth of the patient and parents, number of siblings, number of children and other such data.

As used herein, compensatory mutations are mutations that act in concert with active site mutations by compensating for functional deficits caused by changes or mutations that affect binding in the active site.

As used herein, a relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Such databases are readily available commercially, for example, from Oracle, IBM, Microsoft, Sybase, Computer Associates, SAP, or multiple other vendors.

As used herein, a phenotype refers to a set of parameters that includes any distinguishable trait of an organism. A phenotype can be physical traits and can be, in instances in which the subject is an animal, a mental trait, such as emotional traits. Some phenotypes can be determined by observation elicited by questionnaires or by referring to prior medical and other records. For purposes herein, a phenotype is a parameter around which the database can be sorted.

As used herein, genotype refers to a specific gene or totality of genetic information in a specific cell or organism.

As used herein, haplotype refers refers to two or more polymorphism located on a single DNA strand. Hence, haplotyping refers to identification of two or more polymorphisms on a single DNA strand. Haplotypes can be indicative of a phenotype.

As used herein, a parameter is any input data that will serve as a basis for sorting the database. These parameters will include phenotypic traits, medical histories, family histories and any other such information elicited from a subject or observed about the subject. A parameter may describe the subject, some historical or current environmental or social influence experienced by the subject, or a condition or environmental influence on someone related to the subject. Paramaters include, but are not limited to, any of those described herein, and known to those of skill in the art.

As used herein, computational phenotyping, refers to computer-based processes that assess the phenotype resulting from a particular genotype. The phenotype describes observables, such as, but are not limited to, the structure of the encoded protein, its functional morphological and structural attributes. In particular, as contemplated herein, the phenotype that is assesed is the interaction of a protein with a particular compounds, particularly a drug. As exemplified herein, the

method provides a means to select an effective drug for a particular subjects, particularly mammals, or class thereof.

As used herein, a database refers to a collection of data; in this case data relating to polymorphic variants. Hence a database contains the nucleic acid sequences encoding the variants, or a portion of the variant, such as a portion contianing the active site or targetted site. Additionally, the database may contain other information related to each entry, including but are not limited to, the corresponding 3-D structure of the encoded protein (or a portion thereof) and information regaring the source of each sequence. Some of the entries in a database may be identical, and for purposes herein, a database contains at least 2 different entries, typically far more than 2 entries. The number of entries depends upon the protein of interest and variety and number of polymorphisms that exist. Generally a database will have at least 10 different entries, typically more than 100, more than 500, more than 1000, more than 2000, 3000, 4000, 5000, 8000, 10,000, 50,000, 100,000 and greater. Databases herein containing 20,000 entries and more have been generated and are exemplified herein.

As used herein, a relational database stores information in a form representative of matrices, such as two-dimensional tables, including rows and columns of data, or higher dimensional matrices. For example, in one embodiment, the relational database has separate tables each with a parameter. The tables are linked with a record number, which also acts as an index. The database can be searched or sorted by using data in the tables and is stored in any suitable storage medium, such as floppy disk, CD rom disk, hard drive or other suitable medium.

As used herein, a profile refers to information relating to, but not limited to and not necessarily including all of, age, sex, ethnicity, disease

history, family history, phenotypic characteristics, such as height and weight and other relevant parameters.

As used herein, a biopolymer includes, but is not limited to, nucleic acid, proteins, polysaccharides, lipids and other macromolecules. Nucleic acids include DNA, RNA, and fragments thereof. Nucleic acids may be derived from genomic DNA, RNA, mitochondrial nucleic acid, chloroplast nucleic acid and other organelles with separate genetic material.

As used herein, a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence. By the term "substantially homologous" is meant having at least 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

As used herein, a receptor refers to a molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or synthetic molecules. Receptors may also be referred to in the art as anti-ligands. As used herein, the terms, receptor and anti-ligand are interchangeable. Receptors can be used in their unaltered state or as aggregates with other species. Receptors may be attached, covalently or noncovalently, or in physical contact with, to a binding member, either directly or indirectly via a specific binding substance or linker. Examples of receptors, include, but are not limited to: antibodies, cell membrane receptors surface receptors and internalizing receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles.

Examples of receptors and applications using such receptors, include but are not restricted to:

a) enzymes: specific transport proteins or enzymes essential to survival of microorganisms, which could serve as targets for antibiotic (ligand) selection;

b) antibodies: identification of a ligand-binding site on the antibody molecule that combines with the epitope of an antigen of interest may be investigated; determination of a sequence that mimics an antigenic epitope may lead to the development of vaccines of which the immunogen is based on one or more of such sequences or lead to the development of related diagnostic agents or compounds useful in therapeutic treatments such as for auto-immune diseases;

c) nucleic acids: identification of ligand, such as protein or RNA, binding sites;

d) catalytic polypeptides: polymers, preferably polypeptides, that are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products; such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, in which the functionality is capable of chemically modifying the bound reactant (see, *e.g.*, U.S. Patent No. 5,215,899);

e) hormone receptors: determination of the ligands that bind with high affinity to a receptor is useful in the development of hormone replacement therapies; for example, identification of ligands that bind to such receptors may lead to the development of drugs to control blood pressure; and

f) opiate receptors: determination of ligands that bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

As used herein, prion refers to an infectious pathogen that causes central nervous system spongiform encephalopathies in humans and animals. No nucleic acid component is necessary for the infectivity of prion protein (see, *e.g.*, U.S. Patent No. 5,808,969).

As used herein, a ligand is a molecule that is specifically recognized by a particular receptor. Examples of ligands, include, but are not limited to, agonists and antagonists for cell membrane receptors, toxins and venoms, viral epitopes, hormones (*e.g.*, steroids), hormone receptors, opiates, peptides, enzymes, enzyme substrates, cofactors, drugs, lectins, sugars, oligonucleotides, nucleic acids, oligosaccharides, proteins, and monoclonal antibodies.

As used herein, complementary refers to the topological compatibility or matching together of interacting surfaces of a ligand molecule and its receptor. Thus, the receptor and its ligand can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other.

As used herein, a ligand-receptor pair or complex formed when two macromolecules have combined through molecular recognition to form a complex.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing

a value of 1 for identities and 0 for non-identities) and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps.

Whether any two nucleic acid molecules have nucleotide sequences that are at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA 85*:2444 (1988). Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine identity

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to

Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al., Nucleic Acids Research 12(I)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al., J Molec Biol 215*:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to a reference polypeptide. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

As used herein, AMBER is a force field well known in the arts and designed for the study of proteins and nucleic acids as defined in Weiner et al. J. Comput. Chem. (1986) 7:230-252, where a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (version

3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy. AMBER is available in commercially available molecular modeling programs such as, but not limited to, Macromodel (Columbia University).

As used herein, ECEPP (Empirical Conformational Energies of Peptides Program) is a force field well know in the arts (US Patent No. 5,910,478; 5,846,763). ECEPP/3 refers to version 3 of this well known force field.

As used herein, QSAR refers to structure-activity relationship.

As used herein, vdw refers to van der Waals.

As used herein, RMSD refers to root mean-squared deviation.

As used herein, medical history refers to the parameters and data typically obtained by a physician when examining a subject or other such professional when examining other mammals, and includes such information as prior diseases, age, weight, height, sex and other information. For purposes, the subjects that serve as the source of the samples from which nucleic acids encoding polymorphisms are isolated, include animals, plants, pathogens and any organism that has nucleic acid that exhibits polymorphism. In this context medical history refers to information pertinent to the particular organism.

As used herein, subject history, refers to data such as locale in which the subject was born, raised or resident or visited, and parental history and other such information.

As used herein, a drug is an agent that binds to or interacts with a targeted protein. For purposes, a therapeutic agent is a drug.

**B. Computer-based methods of drug design based on genetic polymorphisms**

Methods for computer-based drug design based on genetic polymorphisms are provided. The methods includes the steps of obtaining one or more, preferably two or more, amino acid sequences of a target

protein that is the product of a gene exhibiting genetic polymorphisms; generating 3-dimensional (3-D) protein structural variant models of all or a portion of the protein from the sequences; and based upon the structures of the 3-D models, designing drug candidates or modifying existing drugs based on the predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants or portions thereof by computationally docking drug molecules with the target protein models; and then, optionally energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity.

A variety of methods that include these steps are provided. Such methods have particularl application, for example, in predicting patient responses. As noted, patients exhibit variable responses to drugs. For some patients a drug may be very beneficial and achieve a desired response; whereas for other patients, with the same disorder, the same drug will have little or no effect. It is known that individuals as well as groups of individuals exhibit a variety of genetic polymorphisms. As described herein, the presence or absence of such polymorphisms can be correlated with the variability of patient responses to drugs.

It is shown herein that by understanding how genetic polymorphisms affect 3-D protein structure of a drug target, for example, it is possible to ascertain the interaction of a particular drug with the target in a particular patient or groups of patients. Based upon this interaction, the outcome can be predicted. It will be possible to determine whether a patient will benefit from a drug or be at risk for a particular side effect. It is possible to predict these responses before exposure to the drug. These

methods also permit rational design of drugs that can treat various populations or ultimately even individuals. These differences and effects can also be taken into account to design drugs that are not dependent upon a particular polymorphism.

Hence, the knowledge derived from understanding the effects of genetic polymorphisms can be used to develop and apply therapeutics more effectively, make clinical trials more successful, for example, by permitting selection of test subjects with the same polymorphism or with polymorphisms for which the drug is designed to interact effectively.

It is shown herein that it is advantageous to use 3-D molecular structures in drug design rather than to consider primary sequence alone. For example, most drugs target proteins either in the afflicted organism or in a pathogen. Disease, drug action and toxicity are all manifested at the protein level. Although the nucleotide sequences of genetic polymorphisms might appear to be quite different, the resulting protein targets may have similar shapes and, therefore, the protein biological function might be the same. Conversely, although genetic polymorphism sequences might appear similar, the resulting proteins may have critical differences in their 3-D structures that greatly affect biological activity. Thus, use of 3-D protein structure models in such methods provide advantages not heretofor realized. Methods for generating 3-D structures are known to those of skill in the art and are also provided herein.

Once the protein target structural models have been selected, structure-based drug discovery methodologies, for example, computational screening or docking programs and methods (*e.g.*, DOCK (available from University of Ca, San Francisco; and AUTODOCK available from Scripps Research Institute, La Jolla), are used to design biologically-active compounds based on the 3-D structures of the biomolecular receptors. Using these methods, drug designers can identify and

computationally rank the various potential clinical drug candidates for maximum efficacy, thereby performing drug discovery *in silico* and avoiding the tedious time and expense associated with *in vitro* drug discovery methods.

In addition to drug design applications, the information derived from studying the structures of biological targets can be used to understand and predict biological responses in patients, such as efficacy, toxicity, drug resistance and other pharmacological effects. Since human clinical trials may cost upwards of $100-300 million, it is desirable to predict the outcome to the greatest extent possible for each prospective drug candidate so that the best prospective drug candidates are advanced to clinical trials. As described below, methods are provided herein for selecting populations for clinical trials.

**1. Methods for obtaining amino acid sequences of a target protein**

Any protein or gene or encoded mRNA that exhibits polymorphisms, herein referred to as the target protein, in structure is contemplated for use herein and for generating the databases as provided herein. The target protein is a protein, polypeptide, or oligopeptide that includes, but is not limited to, receptors, enzymes, hormones, prions, or any such compound with which drugs or other ligands, such as small molecules, peptide agonists, peptide antagonists, other proteins, nucleic acids and other biomacromolecules, interact to bring about a biological response. These target proteins occur in any organism, including plants and animals, eukaryotes and prokaryotes, including pathogens, such as protozoans, parasites, viruses, includind DNA and retroviruses, and bacteria. The protein or gene can be one expressed in the organism, such as molecule targeted for drug interaction , or one expressed in a pathogen.

The target gene is one that exhibits polymorphisms (i.e., sequence variations among a population) and the target protein is the product of a gene exhibiting genetic polymorphisms, or sequence variations, as described herein. Any gene or protein that exhibits polymorphisms is contemplated herein. In particular, genes that encode proteins, polypeptides, or oligopeptides that are targets for drug interaction are contemplated herein. The genetic polymorphisms can occur in the genes of pathogens (*e.g.* viruses, bacteriae, and fungi), parasites, plants, animals, and humans. As such, the sequence a target protein can be obtained by the isolation and analysis of the gene or gene product in samples taken from pathogens, parasites, plants, animals, and humans, most preferably from humans.

The genes or proteins may be isolated from any source, such as animal or plant specimens, or the sequences obtained from any source, including known databases. If starting with gene sequences that include single or multiple nucleotide polymorphisms, the amino acid sequences of the translated proteins can be determined. Protein isolation and sequencing methods are well known to those of skill in the art. Alternatively, samples of the target protein can be obtained and sequenced directly from specimens. Multiple sequence analyses can be performed to determine the exact amino acid variations or mutations resulting from the genetic polymorphisms.

Amino acid sequences of target proteins can also be obtained from data banks and databases (e.g. GenBank, Swiss Prot, PIR) and from publications and other sources in which numerous polymorphisms have been identified and mapped. Samples may be obtained from, for example blood and tissue banks, nucleic acid isolated, genes selected or identified and polymorphims can be mapped from such samples.

2. **Generation of 3-D protein structural variant models**

After the amino acid sequences of target proteins are obtained via the means described in section 1, the 3-D structural models of the sequences of native proteins or of the protein structural variants are then determined. They can be determinedthrough experimental methods, such as x-ray crystallography and NMR, and from structure databases, such as the Protein Databank (PDB). Moreover, 3-D structural models can be determined by using any of a number of well known techniques for predicting protein structures from primary sequences (e.g. SYBYL (Tripos Associated, St. Louis, Mo.), *de novo* protein structure design programs (e.g. MODELER (MSI, Inc., San Diego, CA) and MOE (Chemical Computing Group, Montreal Canada) and *ab initio* methods, see, *e.g.*, U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895), homology modeling, and *ab initio* computational analysis. Homology modeling, structure determination based upon x-ray crystallographic structures, and *ab initio* techniques and combinations of these methods are among those preferred herein.

a. **Homology Modeling**

Homology modeling is based on the relationship between protein evolutionary origin, function and folding patterns. Proteins of related origin and function have conserved sequences and structural features among the members of a homologous family. Using these relationships, a three-dimensional structural model for a protein of unknown structure can be constructed by using composite parts of related proteins in the same family. Where only the primary amino acid sequence of a target protein is known, the sequence can be compared to the sequences of related proteins with known structures (reference proteins), and a model can be built by incorporating the structural attributes of the reference protein together with the sequence of the target protein.

Sequence homology calculations generally require: the amino acid sequence of the target protein; a high resolution structure for at least one, but preferably more, related reference proteins; and any other related amino acid sequences. The reference proteins include structures which are similar to the target protein, either by sequence, fold, function, or which are polymorphisms of the target protein. The more related protein structures and sequences that are available or determined, the more reliable the technique will be at providing an accurate model.

In constructing a protein model using homology modeling, sequence alignment is performed between the target sequence and any known structures within the protein family. Sequence alignment requires determining the similarity between protein sequences by maximizing the number of matches between the sequences while introducing the minimum number of insertions and deletions. Sequence alignment algorithms are well known in the art, and standard gap penalties (*i.e.*, programs that automatically introduce gaps to maximize alignment and then adjust the percentage of identity by applying penalties for gap number and gap length) and other parameters can be selected by the skilled artisan. Additionally, the 3-D structures of the known reference proteins, preferably, are aligned to give the best overall fit for the proteins in the family. This provides indication of structurally-conserved regions, such as regions of the proteins that do not contain insertions or deletions, among the reference structures.

Once the sequences are aligned and the structurally-conserved regions are identified, the coordinates of the reference proteins can be used to construct a 3-D model of the target structure. Coordinates from the protein backbone of the reference proteins are then used to construct the backbone framework for the target protein structure. Side chains can be constructed, for example, by using side chain coordinates from the

reference proteins, searching from a database to obtain side chain conformations that fit in with the existing structural framework or by generating side chains *ab initio* to establish energetically favorable side chain conformations.

The non-conserved regions of the unknown protein can be constructed, for example, using database searching. A database of known protein structures (*e.g.*, PDB) can be searched to identify variable regions in other proteins that have a high degree of sequence similarity to the target sequence and that fit onto the existing structural framework of the protein model. Algorithms for performing sequence similarity matching and homology model building are well known in the art and are available commercially (available from Molecular Simulations, Inc., Tripos, Inc. and from numerous academic sources).

The variable regions can also be modeled by fitting the target sequence to a peptide backbone generated by varying phi and psi angles (*e.g.*, by calculating Ramachandran or Balasubramanian plots, see, Balasubramanian (1974) "New type of representation for Mapping Chain Folding in Protein Molecules," *Nature 266*:856-857) or Balaji plots, see, U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895) of the amino acids to give a loop structure that can be integrated into the model structure based on a sterically and energetically reasonable fit (Figure 1).

In a Balasubramanian plot, the peptide is depicted as a series of different vertical lines, each having solid dots and open circles aligned with the corresponding $\phi$, $\psi$ angle values on the vertical axis, and where each line corresponds to the particular number of the residue having the plotted $\phi$, $\psi$ angles as indicated on a horizontal axis. In the Balaji plot, the values of the $\phi$, $\psi$ angles are shown as the base and tip of a vertical wedge (assuming a vertical angular axis), respectively, with a separate wedge being horizontally positioned on the plot as a function of the

residue number of the $\phi$, $\psi$ angles plotted. The Balaji plot replaces the solid dots and open circles of the Balasubramanian Plot with the base of a wedge and the tip of a wedge, respectively; and further replaces the vertical line joining the dots and open circles of the Balasubramanian plot with the body of the wedge.

**b. Ab initio generation of 3-D structures**

Alternatively, *ab initio* methods can be used in combination with an existing partial homologous structure to generate unresolved portions of the target structure. Such methods are described, for example, in U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895, which as all patents, applications and publications referenced herein, are each incorporated in their entirety. These methods involve: simulating a real-size primary structure of a polypeptide in a solvent box, *i.e.*, an aqueous environment; shrinking the size of the peptide isobarically and isothermally; and expanding the peptide to its real size in selected time periods, while measuring the energy state and coordinates, *i.e.*, the bonds, angles and torsions of the expanding molecule. As the peptide expands to its full size, it assumes a stable tertiary structure. In most cases, due to the manner in which the expansion occurs, this tertiary structure will be either the most probable structure (*i.e.*, it will represent a global minimum for the structure) or one of the most probable structures. The energy equations used to perform the *ab initio* simulation are based on the potential energy of the simulated molecule as described using molecular mechanics.

Once a model is built, it can be refined using energy minimization, molecular dynamics calculations, or simulated annealing as described herein. The steric and energetic quality of the structural models is then evaluated by analyzing the structural attributes of the model, such as phi and psi angles (*e.g.*, by calculating Ramachandran or Balasubramanian or

Balaji plots), or the energetics of the model, such as by calculating energy per residue or strain energy. If the overall quality of the model is not satisfactory, further iterative energy refinement can be performed until the model is considered to be acceptable (*i.e.*, $e_{av} < 1.5$, see below).

A preferred method for generating and refining the structural variant models is illustrated in **FIG. 1**. First, at block 100 of FIG. 1, protein sequence information, derived genetic polymorphisms, is obtained from the methods described earlier. At block 102, the protein is assigned to a protein superfamily in order to identify related proteins to be used as templates to construct a 3-D model of the protein. If the superfamily is not known, sequence analysis or structural similarity searches can be performed to identify related proteins for use as templates in homology modeling studies, as described herein, as indicated at block 104.

Once the conserved regions of the model are assembled, *ab initio* loop prediction (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573) indicated at 106A or *ab initio* secondary structure generation techniques of block 106B, techniques in which the alignments are adjusted using information on the secondary structure, functional residues, and disulfide bonds as described herein, can be used to complete the model (e.g. U.S. Patents Nos. 5,331,573; 5,579,250; and 5,612,895). This model, complete with loops, is then subjected to refinement procedures (block 110) based on molecular mechanics, molecular dynamics, and simulated annealing methods. Energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using, for example, an ECEPP type forcefield (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573) or through molecular dynamics simulations using, for example, a modified AMBER type forcefield (Ramnarayan *et al.* (1990) *J. Chem. Phys. 92:7057-7076*. As known to those of skill in the art a modified AMBER (version 3.3) force field is a fully vectorized version of

AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (see, *e.g.*, Weiner *et al.* (1986) *J. Comp. Chem. 7*:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (*e.g.*,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

The refinement process step 110 is used to offset problems that may arise when homology models are not built carefully or when they are built using fully automated methods. Problems that may arise include chain breaks (e.g. consecutive $C^{\alpha}$ atoms are farther apart than the optimum distance of 3.7 to 3.9 Å); distorted geometry (e.g. bond lengths and bond angles are too far from their optimal values); *cis*-peptide bonds (*e.g.*, incorrect isomerization of the peptide backbone in non-proline residues when it is not required); disallowed backbone and side-chain conformations (*e.g.*, dihedral angles do not satisfy the Ramachandran plot (see, Balasubramanian (1974) *Nature 266*:856-857) criteria for a fully favorable protein structure conformation); and misfolded loops (e.g. non-homologous loops are generated in unnatural conformations). The refinement procedure 110 removes distortions of covalent geometry by using energetic methdods, converts disallowed backbone and side-chain conformations into allowed ones using simulated annealing methods, conserves protein core structure and secondary structural elements built by homology, and rebuilds unnatural loop constructions (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573).

For quality control (block 112), the protein structural characteristics, for example, stereochemistry (*e.g.*,, phi/psi and side chain angles), energetics (*e.g.*,, strain energy), packing profile (*e.g.*,, packing factor per residue) and hydrophobic packing are evaluated and required to meet acceptable criteria before the structures are used in further studies

or inputted into a structural polymorphism database. Quality control using strain energies entails computing normalized residue energies (NREs) based on the equation:

$e_i = [E(i,X) - E_{AV}(X)] / E_{SD}(X)$, where

E(i,X) is the energy of interactions of amino acid X in position i with protein environment and solvent;

$E_{AV}(X)$, $E_{SD}(X)$ is the average residue energies and their standard deviations calculated for 20 amino acids in more than 100 high-quality crystal structures; and

NREs characterize how favorable the interactions of each residue are within the protein environment (Majorov and Abagyan, (1998) Folding & Design 3:259).

The average NRE characterizes the overall quality of a protein structure and is defined as:

$e_{av} = (1/N) \Sigma_i e_i$, where

$e_{av} \leq 0.5$ denotes high-resolution X-ray crystal structures;

$e_{av} \leq 1.0$ denotes good as NMR and theoretical models; and

$e_{av} \geq 1.5$ denotes structures that require further refinement.

After the quality of structure is determined at block 112, the model is checked at block 114 to determine if it is satisfactory. If the overall quality of the model is not satisfactory, a "No" outcome at block 116, then remedial action is undertaken to fix problems at block 118, including further iterative energy refinement (block 110), and repeated checking (block 114). The refinement and evaluation is repeated until the model is considered to be acceptable, a "Yes" outcome at block 120, whereupon structural and/or physical properties (e.g. energetics and phi/psi angles) are calculated at block 122A and clinical data (if available) is obtained at block 122B. The model is then inputted into a structural polymorphism database at block 124.

FIG. 2 shows an exemplary method for generating structural variant models derived from genetic polymorphisms and using them in structure-based drug design studies. At the block numbered 200, patient data is acquired for a gene that exhibits genetic polymorphisms. Protein sequence information is then derived, at block 202. A check is made for determination of the 3-D structure of the native protein. If the 3-D structure has been determined, a "Yes" outcome at block 206, then a multiple sequence analysis is performed at block 208 to determine the exact amino acid variations for the structure. If the 3-D structure has not been determined, a "No" outcome at block 210, then the structure is determined using physiochemical methods at block 212.

Next, at block 214, the 3-D structural models for all variants are generated. A refinement process is then completed at block 216 for the structural models. As noted above in connection with FIG. 1, the process involves subjecting each model, complete with loops, to refinement procedures based on molecular mechanics, molecular dynamics, and simulated annealing methods. As before, the energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using an ECEPP type forcefield (Dudek *et al.* (1998) *J. Comp. Chem. 19*:548-573), or through molecular dynamics simulations using, for example, a modified AMBER type forcefield (Ramnarayan *et al.* (1990) *J. Chem. Phys. 92*:7057-7076), where a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (Weiner et al. (1986), J. Comp. Chem. 7:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (*e.g.*,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

At block 218, a quality evaluation is performed for all the models. As described in connection with the quality evaluation process in Fig. 1, the evaluation at block 218 involves evaluating the protein structural characteristics, for example, stereochemistry (*e.g.*, phi/psi and side chain angles), energetics (*e.g.*, strain energy), packing profile (*e.g.*, packing factor per residue) and hydrophobic packing, which must meet acceptable criteria before the structures are used in further studies or inputted into a structural polymorphism database.

After the model quality is determined, at block 220 the models are checked to determine if they are satisfactory for further use. If a model is not satisfactory, a "No" outcome at block 222, then the problems are identified and solved with remedial action at block 224. The remedial action may include further iterative energy refinement at block 216 and repeated checks of model quality at block 218. Once the models are satisfactory, a "Yes" outcome at block 226, structure-based drug design methods are applied at block 228 to identify potential new drugs that bind to the structural variant models. The drug design methods are described further below.

FIG. 3 shows another exemplary and alternative method for generating structural variant models derived from genetic polymorphisms and using them in structure-based drug design studies. The process of FIG. 3 is similar to the process of FIG. 2 from the initial process at block 300 of acquiring patient data for a gene that exhibits genetic polymorphisms through the process of obtaining models that are satisfactory (a "Yes" outcome at block 326). Thus, block numbers in FIG. 3 from 300 through 326 that correspond to FIG. 2 blocks numbered from 200 thorough 226 refer to similar operations. Unlike FIG. 2, however, the process illustrated in FIG. 3 then involves docking operations.

At block 328, once the models are determined to be satisfactory, drug molecules are docked with the structural variant models. Next, at block 330, the free energy of binding is evaluated with the potential drugs under study for each structural variant model. At block 332, the total free energy of binding is decomposed, based on the interacting residue in the protein active site. Lastly, at block 334, the free energy of binding is correlated with patient data, if the data is available. Thus, the 3-D structural data is employed in drug design. Details of using such structural data in drug design are described further below.

**c. Crystal structures**

The crystal structure of any protein can be determined empirically and the resulting coordinates used as the basis for determing structures of variants. Such structures are often known (see, *e.g.*, Kohlstaedt *et al.* (1992) *Science 256*:1773-1790 for a crystal structure of HIV-1 RT bound to a ligand).

**3. Use of 3-D structural variant models in drug design**

The structural differences in protein structural variants that arise due to genetic polymorphisms can have profound effects on biological activity. Because of the structural differences among the variants, they may have different physical or reactive properties and therefore may exhibit different biological activities. These differences may include, for example, different responses to a given drug, so that a drug which works well in a patient with one particular genetic polymorphism may not work as well in another patient exhibiting a different polymorphism.

The 3-D molecular structures of drug targets derived from genetic polymorphisms can be used in structure-based drug design studies to greatly advance the development of new pharmaceuticals. Relational databases of these 3-D structures that are derived from samplings of genetic polymorphisms over a patient population or a cross-section of the

population can be used to design potential drugs in order to optimize effectiveness for the particular population.

The structures and databases described herein can provide information that is useful, for example, in designing a drug that is effective in the greatest percentage of the population. It is desirable that a given drug is effective in the largest percentage of the population, since such a drug is likely to have the greatest clinical utility and thus the greatest commercial value. A drug with superior performance properties is sometimes referred to as a "best in class" drug and is highly prized by pharmaceutical companies since this heralds market leadership and the likelihood of commercial success. The databases and methods described herein can be used to determine 3-D protein structures for drug targets that are associated with particular genetic polymorphisms and to use the structures in drug design studies for design and optimization of candidate drugs that exhibit activity over the broadest patient population.

Genetic polymorphisms may result in target protein structural variants in which drug efficacy correlates with specific populations or subpopulations. In some cases, it might be desirable to target drug design or drug therapy toward a specific patient population, such as a particular race, gender, or age group, affected by a certain disease or condition or toward those having a specific genetic polymorphism. The information derived from comparing the 3-D structural variants arising from different genetic polymorphisms may be useful for understanding why drugs are active or inactive in different subpopulations, or for assisting in developing new drugs to maximize efficacy across specific populations.

a. **Selection of relevant structural variants**

The structural variant models in the structural polymorphism database provided herein can be used to design new drugs or to select a drug therapy that would be appropriate for a patient exhibiting a particular genetic polymorphism. As it may not be possible for a drug to work equally well for all polymorphisms, and thus all patients, representative structural variants can be selected for use in drug design studies in order to maximize biological activity based on genetic polymorphisms.

In some cases, structural variants are analyzed to determine the common structural features that are conserved through the selected models. These conserved features are used as a basis for drug design. In some cases, the structural variant corresponding to the genetic polymorphism occurring most commonly in a population can be selected for use in identifying drugs that would be effective in the greatest percentage of the population. Optionally, structural variants corresponding to a relevant subpopulation, such as a particular gender, age, race, or other characteristic, can be selected for use in designing drugs that are active in that subpopulation. In other cases, individual structural variant models can be selected for use in designing drugs that are specifically active against one target in one individual arising from a particular genetic polymorphism. Additionally, model structures that represent variants derived from patients that receive a specific treatment regimen or exhibit a particular clinical response (e.g. drug resistance) to a given drug are used as bases for drug design.

The relevant structural variants may be identified using the structural analysis tools described herein, optionally in combination with database and statistical analysis tools that permit a complete analysis and comparison of the molecular structures and properties of the structural

variants. The structural variants selected based on the criteria including, but not limited to, those listed above are used in drug design.

**b. Drug design**

Once the protein target structural models have been selected, structure-based drug discovery methodologies, for example, computational screening or docking (*e.g.*, DOCK (available from University of Ca, San Francisco; and AUTODOCK available from Scripps Research Institute, La Jolla and others referenced herein or known to those of skill in the art), can then be used to design biologically-active compounds based on the 3-D structures of the biomolecular receptors.

Using these methods, drug designers can identify and computationally rank various potential clinical drug candidates for maximum efficacy, thus cutting the time and expense associated with drug discovery. The preferred design of drug candidates or the modification of existing drugs is based on the intermolecular interactions between the drug candidate or modified drugs and the selected structural variants predicted by computationally docking drug molecules with the target protein models; energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity.

**c. Computational docking**

Methods for using the structural variant models to design potential new drugs or to aid in the selection of a drug therapy based on the interactions of selected small molecules with the particular variants are provided. Structure-based drug design experiments, such as computational screening or docking studies, calculation of binding

energies or analysis of steric, electrostatic or hydrophobic properties of the resulting structural variant models, can be performed on selected structural variant models to aid in the understanding of observed biological activities or to determine new potential drug candidates to bind to the particular target.

In a typical computational docking protocol, the active site, or sites deemed important for protein activity, of the protein model is defined. A molecular database, such as the Available Chemicals Directory (ACD) or any database of molecules, is screened for molecules that complement the protein model. Solvation parameters are factored in (see, *e.g.*, Shoichet *et al.* (1999) PROTEINS: Structure, Function, and Genetics 34:4-16). In these computational docking studies, drugs or drug candidates are fitted to the structural variant models based on complementary interactions (*e.g.*, steric, hydrophobic, or electrostatic interactions). Methods for performing such studies are well known and software tools for performing the calculations are widely available (M. Lambert, "Docking Conformationally Flexible Molecules into Protein Binding Sites" in Practical Application of Computer-Aided Drug Design, Charifson, Ed., Marcel Dekker, NY, pp. 243-303; Kurtz (1992) *Science 257*:1078-1082; Kuntz et al. (1982) J. Mol. Biol. 161:269-288; Stewart *et al.* (1992) *Med. Chem. Res. 1*:439-443; Shoichet *et al.* (1993) *Science 259*:1445-1450; Shoichet *et al.* (1991) *J. Mol. Biol. 221*:327-346).

New potential drug candidates can be designed by identifying potential small molecule drugs that can bind to a particular structural variant. This is accomplished, for example, by methods including, but are not limited to, methods for electronic screening of small molecule databases as described herein, methods involving modifying the functional groups of existing drugs *in silico*, methods of *de novo* ligand design. Methods for computationally desiging drugs are known to those

of skill in the art and include, but are not limited to, DOCK (Kuntz *et al.* (1982) "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288; available from University of Ca, San Francisco); and AUTODOCK (see, Goodsell *et al.* (1990) "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202; available from Scripps Research Institute, La Jolla); GRID (Oxford University, Oxford, UK); CAVEAT (UC Berkeley, Ca), LEGEND (Molecular Simulations, Inc., San Diego, CA); LUDI (Molecular Simulations, Inc., San Diego, CA); HOOK (Molecular Simulations, Inc., San Diego, CA); CLIX (CSIRO, Australia); GROW (Upjohn Laboratories, Kalamazoo); others including HINT, LUDI, NEWLEAD, HOOK, PRO-LIGAND and CONCERTS (see, M. Murcko, "An Introduction to De Novo Ligand Design" in Practical Application of Computer-Aided Drug Design, Charifson, Ed., Marcel Dekker, NY, pp 305-354), methods based on QSAR (quantitative structure-activity relationships, *QSAR and Drug Design: New Developments and Applications*, Fugita, Ed., (1995) Elsevier, pp 3-81; 3D QSAR in Drug Design, Kubinyi, Ed., (1993) Escom, Leiden), and other methods known to those of skill in the art for determining molecules that have optimal binding interactions with a selected target.

The docked complexes, if needed, are further refined energetically to optimize geometries within the binding site and to select the best structure from a set of possible structures, using molecular mechanics, molecular dynamics, and simulated annealing techniques, including those described herein and others that are known to those skilled in the art.

**d. Free energy of binding studies**

After the computational docking step, the free energy of binding of the docked complex is calculated, and the total free enegy of binding is decomposed based on the interacting residues in the protein active site or

sites deemed improtant for protein activity. Analyses of the binding energies are needed to identity drug candidates. If need or desired, the free energy of binding of different drugs or potential drugs to each structural variant model can be calculated by substracting the free energy of the non-interacting protein and drug from the free energy of the protein-drug complex. The total free energy of binding is decomposed into its various thermodynamic components, e.g. enthalpic and entropic components, based on the interacting residues in the protein active site in a solvated model to characterize the structural and thermodynamic features in the mode of drug binding and to determine the contribution of the solvent] (see, *e.g.*, Wang *et al.* (1996) *J. Am. Chem. Soc. 118*:995-1001; Wang *et al.* (1995) *J. Mol. Biol. 253*:473-492; Ortiz *et al.* (1995) *J. Med. Chem. 38*:2681-2691, which describes a computational method for deducing QSARs from ligand-macromolecule complexes). Following the computational drug design protocol described herein, any potential new drugs that are identified can be synthesized in, for example, industry or academia, and subjected to further biological testing, such as *in vitro* studies or pre-clinical and clinical *in vivo* testing.

Based on the predicted intermolecular interactions of the drugs or modified drugs with the structural variant models from binding studies, potential drug candidates that are specific for a protein with a selected polymorphism or that specifically interact with all proteins exhibiting the polymorphism can be identified.

It is also possible to individualize drug design or drug therapy by determining the structural variants associated with a particular patient and then designing or screening drugs or potential drugs to maximize efficacy in that subject or in a subpopulation that exhibits the same genetic polymorphism. The variants may also be used to track polymorphic variations in infectious organisms, such as viruses. For example, the

human immunodeficiency viruses (HIVs) reverse transcriptase and protease have served as drug targets (see, Erickson *et al.* (1996) *Ann. Rev. Pharmacol. Toxicol 36*:545-571); their three-dimensional structures are known (see, *e.g.*, Nanni *et al.* (1993) *Perspectives in Drug Discovery and Design 1*:129-150; Kroeger *et al.* (1997) *Protein Eng. 10*:1379-1383). The clinical emergence of drug-resistant variants of these viruses has limited the long-term effectiveness of drugs targeted against these enzymes.

As noted, these enzymatic proteins in order to preserve function must exhibit conserved 3-D structures. The methods herein permit design of drugs specific for the conserved regions of the 3-D structures. They also permit selection of drug regimens based upon the alleles expressed. Hence, methods for designing HIV enzyme-specific drugs are provided. Flow charts illustrating exemplary alternative embodiments using protein 3-D structures derived from genetic polymorphisms in structure-based drug design studies are provided (see, Figs. 2 and 3). In the flow charts depicted in these figures, the drug design includes structure-based drug design methods (see, Figure 2) and computational docking of drugs with structural variants, evaluation of the binding energy of the docked complexes, and correlation of the binding energy with patient data such as age, gender, race, drug treatment history, and any other pertinent information that is available (see, Figure 3). The data generated by this computer-based method can be stored in a database, such as, for example, in a relational database. The resulting database can be screened using searching tools to select potential drugs and therapeutic agents that bind to or exhibit biological responses towards target proteins.

C. **Applications of computer-based methods**

As discussed above, the computer-based methods provided herein include some or all of the steps of obtaining one or more, preferably two or more, amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms; generating 3-dimensional (3-D) protein structural variant models from the sequences; and based upon the structures of the 3-D models, designing drug candidates or modifying existing drugs based on the predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants by computationally docking drug molecules with the target protein models; energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity. There are numerous applications of these methods, which include structure-based drug design and drug testing; selection of clinically relevant populations for drug testing and other such methods.

1. **Genetic polymorphisms and structure-based drug design**

As noted above, structure-based drug design is an increasingly useful methodology that has made a great impact in the design of biologically active lead compounds. Drug designers can design and screen potential new drugs via computational methods, such as docking or binding studies, before actually beginning patient testing. The drugs designed by such methods, and also those identified by traditional methods of drug discovery, are then tested in clinical trials. Among those that show efficacy for a particular indication and low toxicity ultimately are approved for use. It is found, however, that not all patients with a

particular indication respond uniformly to the drugs. The drug may not be efficacious or side-effects may be pronounced.

The methods provided herein, represent a further advance in the use of rational drug design methods. As described herein, polymorphic variation has an effect upon the 3-D structure of encoded proteins. As a result, drugs interact with variants differently, leading to differential responses in the population as a whole. A new approach to drug design and testing is provided herein. This methods involves identifying polymorphisms and determining 3-D resulting structures, which are then used in methods, including, computational drug design, in the selection of patient populations, in designing treatment protocols and in other applications.

**2. Drug resistance**

Methods for understanding and overcoming drug resistances by using 3-D protein model structures resulting from multiple genetic polymorphisms or mutations in an infectious agents, such as viruses, bacterial and other pathogenic agents are provided. Also provided are methods that for using this information in drug design studies.

In the case of infectious organisms or other replicating or mutating agents, such as flu, HIV, rhinovirus or biological warfare agents, some polymorphisms or mutations may arise over time which convey resistance or susceptibility to specific drug therapy, for example, by altering the drug target structure or physical properties so that a specific drug or therapy, such as an antibiotic or vaccine, may no longer be able to bind to or otherwise interact with the target protein to exert its desired biological effect. For certain infectious agents, such as HIV, genetic polymorphisms in certain genes give rise to drug resistance as the virus mutates (see, *e.g.*, Erickson et al. (1996) Annu Rev. Pharmacol. Toxicol. 36:545-571).

Where drug resistance that arises from mutations or polymorphisms is observed, the methods described herein can be used to develop new drugs that overcome the resistance. For example, once drug resistance is observed, the structure associated with the resistant polymorphism can be determined and used in further drug design studies to suggest new drugs or modifications to the existing drug that will restore biological activity by targeting different mutants or that will target multiple mutants simultaneously.

The model structures can also be used to correlate drug resistance in infectious diseases with the structural variants derived from genetic polymorphisms. Here, the 3-D structure of the virus or other drug target is determined for the particular variant model against which the drug was effective. When drug resistance arises due to a genetic polymorphism, a model for the structure variant associated with the resistant organism can be generated, and a new drug can be designed or modifications can be made to the existing drug to overcome the resistance.

For example, samples of the mutating organism can be obtained over time and structural models for the resulting proteins can be generated. These models can then be used to design new drug therapies that are active against the mutated organism. Multiple drug resistant structures can be analyzed to obtain an average structure or to identify common structural features in order to design new drugs that have the broadest spectrum of activity against multiple mutations.

Such structural information is useful in designing effective drug therapies to overcome resistance or to develop drugs that are effective over a range of genetic polymorphisms and thus work for the maximum number of patients.

### 3. Identification of conserved structural features or pharmacophores

If common structural features are observed over a range of protein targets that are derived from genetic polymorphisms, these common features may be used to design a drug that is effective with a variety of genetic polymorphisms and thus many patients. The retention of certain common structural features over a large number of genetic polymorphisms suggests that those features may not be mutatable because the conserved structure may be essential to protein function, *e.g.*, to the viability of an infectious organism or virus. Such conserved structural elements are prime targets for structure-based drug design, *e.g.*, anti-infective or antibiotic drug design, and can lead to highly effective therapies.

The common structural features can serve as a basis for structure-based drug design, for example, by serving as a scaffold for building a receptor model into which potential drug candidates can be docked or as a pharmacophore query for screening a library of physical or virtual chemical or biochemical molecules to identify compounds that match the pharmacophore template and, thus, are potential drug candidates.

Analysis of 3-D protein structural variants derived from genetic polymorphisms to identify the common structural features over a large number of structural variants can aid in the design of drugs that are active over a broad range of genetic polymorphisms, such as in a large number of patients or against drug resistant targets.

In comparing sets of related protein structures, such as those with the same biological function or those resulting from genetic polymorphisms, certain parts of the structural framework are often found to be conserved, while other parts vary among the proteins. Mutations that occur in the conserved regions of the structure can have significant effects biological activity. For example, in viruses, the conserved features

can be essential to protein function and, thus, to the viability of the infectious organism or virus. Identifying the conserved structural features over a range of structures often gives insight into which structural features are necessary for biological activity and are therefore non-mutatable. By analyzing a number of structural variants derived from genetic polymorphisms that exhibit drug resistance, it is possible to identify or design drugs that interact best with the common structural features in all of the variants. Using these features in structure-based drug design studies leads to the identification of drugs that retain biological activity despite multiple mutations, or polymorphisms, and could help to overcome the problem of drug resistance.

In certain preferred embodiments, new potential drug candidates can be identified using the structural variant models by identifying pharmacophores or conserved features in the protein structural variant models and using this structural information to identify small molecules that would bind to the structural variant models.

Using structural comparison tools described herein, the common structural features that are conserved across a range of structural variant models of a given protein based on different genetic polymorphisms can be identified. To do this, multiple structural variant models are compared, generally by superimposing the coordinates of one variant model onto those of one or more other variants and observing the structural fit. Such functionality is commonly found in molecular graphics or homology modeling packages. Once the optimum fit of structures is performed, then the structural features that are present throughout the structural variant models can be identified and used as the basis for drug interactions in structure-based drug design studies. For example, the pharmacophores or conserved features can be specified as database queries and a library or database of small molecule structures can be

searched to identify new lead compounds to bind to the pharmacophores. Alternatively, other structure-based ligand design strategies can be employed to design lead compounds or to identify modifications to be made to existing drugs to improve biological activity.

**4. Identification of compensatory structural changes**

Certain proteins, for example, viral proteins or other infectious organisms, may harbor multiple genetic polymorphisms. Since each genetic polymorphism can give rise to slight changes in structure, some, and over time, many, additional genetic polymorphisms may cause changes in the protein structures that significantly affect biological activity. These structural changes could result in, for example, different dynamical behavior, alteration in enzyme kinetics or differences in substrate recognition, which can significantly alter drug response. For example, a mutation for one drug compound can suppress a mutation to a second drug due to compensatory effects. In these cases, a drug which is predicted to be ineffective for a given patient based upon the single nucleotide correlation may, in fact, be effective as a result of these changes.

Because mutations are so frequent in AIDS and other viruses, few sequences are exactly the same in different patients. Thus, it is difficult or inconclusive to generate multiple mutation sequence correlations for drug resistance. If each patient has a different viral sequence due to a high viral mutation rate, then no sequence correlation is even possible in such cases.

The methods described herein can be used to study the effects of multiple genetic polymorphisms on a resultant protein structure. Multiple mutations are common in AIDS and other viruses, which makes sequence correlation difficult. By observing the structural effects of the mutations on the resulting protein, it is possible to look at the net effect of all

structural changes and to consider the overall structure of the protein in drug design studies. For example, a mutation might occur in the active site, or site of drug action, in a protein. Additionally, there may be related mutations in other parts of the protein structure, which might not be identified from a single point mutation correlation. These related mutations could have an effect on biological activity of the protein. By looking only at the active site, it might be predicted that a drug or potential drug would not bind to the protein. The additional mutation, however, might cause compensatory structural changes in the protein structure that alter its properties in a way that restores biological activity.

By computing 3-D protein structures from gene sequences containing multiple polymorphisms, it is possible to more accurately predict the effect of multiple sequence mutations on protein structure and, thus, to obtain a better correlation between sequence and drug resistance than by considering sequence correlations alone. This information can be useful, for example, in understanding drug resistance and can aid researchers and clinicians in developing new drug therapies to overcome drug resistance.

The structures that are derived based on multiple generic polymorphisms can be used in structure-based drug design studies to provide frameworks, or scaffolds, into which drug or potential drug molecules can be docked. This permits the design of drugs that are active against a wider range of structural variants, thus, in more patients or against a range of drug resistant proteins.

**5. Clinical Applications**

A knowledge of the repertoire of structural differences arising from genetic polymorphisms across the human population or specific subpopulations can provide insight into the differing biological responses in patients based on their genetic differences. For example, where clinical

data are available for patients having particular genetic polymorphisms, this information can be associated with the 3-D protein structural variants and used to find correlations between polymorphisms and observed drug responses.

The methods provided herein can be used to design drug therapies that bring about favorable clinical responses (or eliminate unfavorable effects) in patients, to identify pharmacological effects of drugs in different patient subpopulations (e.g. age, race, gender) and to simulate clinical trails to increase the probability that the trials will yield optimal results.

Because of the high cost of clinical trials, such studies are generally focused on small patient populations. The structural analysis tools described herein permit the extension of clinical trials to cover patient populations not specifically included in the study. This is accomplished through correlation of the structural variants derived from genetic polymorphisms with clinical responses.

The molecular structures and databases described herein can also find application in the understanding and prediction of clinical or pharmacological drug responses, for example, efficacy, toxicity, dose dependencies or side effects in patients. For example, relational databases containing 3-D protein structural variants can provide a means for managing and using the information to understand and predict clinical responses in patients.

In other embodiments, observed clinical data from patients in a clinical trial can be associated with the structural variant models for each genetic polymorphism exhibited in the clinical subjects, for example, in a structural polymorphism relational database. The correlation between the structural variants and observed clinical effects can then be utilized to predict clinical outcomes in patients that did not participate in the clinical

trial. For example, a structural variant model can be generated for a patient based on a genetic polymorphism exhibited in the patient, and the database can be mined to identify structurally similar variants for which clinical results are known. Structural similarity can be determined, for example, by superimposing the structures and measuring the RMS (root mean squared) differences between the structures or by using pattern matching or motif searching algorithms. The results can be used to predict clinical responses in the patient based on the clinical data associated with the structurally similar variants.

The predicted correlations can also be used to aid in the design of subsequent clinical trials. The follow-on trials can be made more effective through the judicious selection of patients with given genotypes (i.e., those exhibiting the same genetic polymorphisms), as guided by the structurally predicted outcomes. For example, a clinical trial can be designed based on a subpopulation of clinical subjects which exhibit a specific genetic polymorphism (i.e. structural variant) to demonstrate the effectiveness of a given therapeutic on a targeted population.

In other embodiments, the methods provided herein can be used in the selection of drug therapies for patients exhibiting a particular genetic polymorphism. This is accomplished by generating the structural variant model associated with the polymorphism, docking drug molecules that might be used to treat the patient into the structural variant model and calculating the binding energies of each drug with the variant. The results of docking or free energy calculations can be correlated to clinical data, for example, patient population (*e.g.*, ethnic background, race, sex, age), treatment regimen, patient response to a particular drug or duration of treatment. The binding energies can be compared, for example, to determine which drug would best bind to the variant in order to identify

the drug that could best be used to treat the patient to optimize biological activity.

**D. Creation of 3-D Structural Polymorphism Databases**

The above-noted methods all rely upon the use of databases of nucleic acid sequences. Any such database known to those of skill in the art may be employed; numerous such databases are publically available (e.g. the Stanford HIV database). The Stanford HIV database is hierarchal database with information about HIV patients who received or did not receive protease inhibitor treatments, patient-dates, isolates, sequences, hyperlinks to MEDLINE and GenBank abstracts, and art. This database, however, does not contain 3-D protein structures of any proteins including HIV reverse transcriptase (RT) and HIV protease (PR; see, *e.g.*, Shafer *et al.* (1999) *Nucleic Acids Res. 27*:348-352, Shafer *et al.* (1999) *J. Virol 73*:6197-6202, http://hivdb.stanford.edu/hiv, Richter (January 20, 1999) "AIDS drugs found to be effective in the world's most common HIV strains).

Databases of sequences and associated information may also be generated as described herein by obtaining samples and sequences from a variety of sources. In all instances, further databases are generated by then calulating 3-D structural models of the encoded proteins or relevant portions, such as active binding sites, thereof, from the nucleic acid sequence information. It is these databases of nucleic acid sequence and/or primary protein sequence and the associated 3-D structure that are provided herein and that are used in the all of the methods, except for the computational phenotyping discussed below, which does not require a database, provided herein. Hence databases comtaining computationally determined 3-D structures of polymorphic proteins or portions thereof are provided herein. These databases serve as tools in a variety of methods, including those provided herein.

Databases that include 3-D structures for variant proteins encoded by the nucleic acids that contain polymorphisms are provided. These are generated after 3-D structural models are constructed for the protein structural variants, preferably for all of the protein structural variants, representing the genetic polymorphisms, by inputting the atomic coordinates into a structural polymorphism database, preferably a relational database, and optionally with associated structural and/or physical properties (*e.g.*, phi/psi and side-chain angles and energetics), and other data, if available, including, but are not limited to, historical data, such as parental medical histories, and clinical data. The resulting database is used in structure-based drug design studies and for clinical analyses. Figure 11 is a tabulation of the 3-D coordinates of a representative entry, an HIV protease, that is encoded by the DNA in one of SEQ ID Nos. 3-74 and 77-117, and that is an entry in an exemplary database that includes 3-D structures. Exemplary databases that contain the nucleic acids sequences and structures of all proteins encoded by SEQ ID Nos. 3-117 as well additional nucleic acids are provided herein and are described in the EXAMPLES.

A database is preferably interfaced to a molecular graphics package that includes 3-D visualization and structural analysis tools, to analyze similarities and variations in the protein structural variant models (see, copending U.S. application Serial No. 09/531,995, which is published as International PCT application No. WO 00/57309, and is a continuation-in-part of U.S. application Serial No. 09/272,814, filed March 19, 1999). Briefly, International PCT application No. WO 00/57309 provides a database and interface for access to 3-D molecular structures and associated properties, which can be used to facilitate the design of potential new therapeutics. The interface also provides access to other structure-based drug discovery tools and to other databases, such as

databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. This interface can be modified as needed to adapt for use with a paritcular database.

A relational database that collects multiple data files relating to the same molecular structure in the same subdirectory and that provides an interface to access all of the collected files from the same structure using the same user interface program is also provided. The collected files include a variety of information and computer file formats, depending on the type of information to be conveyed to users of the database. In practice, a user communicates over a public network, such as the Internet, or over a controlled network, such as an internet, with a secure file server that controls access to the collected files, and the interface to the collected files is provided by a standard graphical user interface program that is widely available. In this way, a convenient means of searching molecular structure data for characteristics of interest is provided. Data searching, file viewing, and investigation of multiple representations of molecular structures from within a single viewing program can also be performed using the database and interface.

The data files can be those available over a wide network such as the Internet, and a suitable graphical user interface designed or obtained. Such interface is used for viewing the data files is a standard Internet web browser program, such as the web browser products by Netscape Communications, Inc. and Microsoft Corporation that are distributed free

of charge. Such browser products readily import and provide views of files having a wide variety of formats that contain alphanumeric, video, and audio data. A security server is preferably located between the user browser program at a network client machine controls access to the database, which is housed at a file server connected to the security server. Before a user gains access to the database, the security server checks authorization for the individual user and then, if appropriate, permits downloading of appropriate data from the database file server. It is contemplated that the databases containing 3-D structures of proteins or portions thereof the exhibit polymorphism will be loaded.

Data for a molecular structure is loaded into the database by specifying the file pathnames for the various data files that contain the different types of data, including the different molecule views. Using a browser to view the data files permits various helper applications, called plug-ins, to smoothly and transparently accept the different file formats and provide views to the user. The various data files of the database are organized in accordance with the database design when they are loaded into the database and are managed by a relational database management program.

In addition to 3-D protein structures and associate primary sequences, as provided herein, the database can optionally contain associated biological or clinical data, such as drug resistance, side effects, efficacy, pharmacokinetics and other data, that correlate with or can be correlated the structural variants. This information will be used for correlating observed clinical effects to specific structural variants and for predicting clinical responses and outcomes based on a patient's structural variants, *i.e.*, genetic polymorphisms.

Structural analysis tools are preferably integrated with the structural database for comparing and analyzing the resulting protein structural variant models. For example, the molecular graphics software package described in International PCT application No. WO 00/57309, includes structural analysis capability to measure the structural attributes of the model (distances, angles, etc.), to analyze sequences and secondary structures, to study physical properties such as hydrophobicity, electrostatic potential, and active or reactive sites in the protein, as well as to evaluate the quality of the structure (both conformationally and energetically).

Structures can also be compared by aligning them, such as by performing a least squares fitting of the x-, y- and z-coordinates of each of the structural variant models and superimposing the structures or any other alignment method or structural comparison method. For example, the structures of the variants can be clustered, or grouped together, based on structural similarity. This can save time over studying each structural variant independently because, where structures are considered to be similar enough that they are clustered together (*e.g.*, if their structures can be superimposed within a specified tolerance), then only a representative structure, or perhaps an average structure or scaffold, which is derived as a composite of the individual structural variant models, can be used in further drug design studies.

Tools for database searching can also be included in the software package. These can be used to query the database for structural variant models having similar properties, such as molecular structure or sequence similarity. These tools are used, for example, to mine the database to identify variant models that are structurally similar (*e.g.* to find structures that overlap within a specified tolerance), and thus would be predicted to interact in the same way with potential drugs or exhibit the same clinical

response. This information could be useful in understanding the structural or clinical effects of different genetic polymorphisms and could potentially save time and money by extending the results of previously performed clinical or computer-based drug design studies to predict the results of studies on similar structural variants that have not yet been performed.

**1. Exemplary Databases**

Databases containing data representative of the 3-D structure of structural variants encoded by a selected gene or genes or the 3-D structure of other polymorphic variants are provided. The selected genes can be drug target, such as receptors and genes of infectious agents, such as the HIV protease or reverse transcriptase. Exemplary databases are presented in Example 5 which describes the construction, interface, use and appliations of HIV PR and RT databases. These databases may be stored on any suitable medium and used in any suitable computer system. Systems and methods for generating, storing and processing databases are well known.

**2. Computer systems**

Computer systems for processing the databases and computer systems containing the databases are provided. The processing that maintains the database and performs the methods and procedures using the databases may be performed on multiple computers, or may be performed by a single, integrated computer. For example, the computer through which data is added to the database may be separate from the computer through which the database is sorted or analyzed, or may be integrated with it. Each computer operates under control of a central processor unit (CPU), such as a "Pentium" microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. A computer user can input commands and data from a

keyboard and display mouse and can view inputs and computer output at a display. The display is typically a video monitor or flat panel display device. The computer also includes a direct access storage device (DASD), such as a fixed hard disk drive. The memory typically includes volatile semiconductor random access memory (RAM). Each computer preferably includes a program product reader that accepts a program product storage device from which the program product reader can read data (and to which it can optionally write data). The program product reader can include, for example, a disk drive, and the program product storage device can comprise removable storage media such as a magnetic floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, or a DVD data disc. If desired, computers can be connected so they can communicate with each other, and with other connected computers, over a network. Each computer can communicate with the other connected computers over the network through a network interface (see, *e.g.*, Examples below) that permits communication over a connection between the network and the computer.

The computer operates under control of programming steps that are temporarily stored in the memory in accordance with conventional computer construction. When the programming steps are executed by the CPU, the pertinent system components perform their respective functions. Thus, the programming steps implement the functionality of the system as described above. The programming steps can be received from the DASD, through the program product reader, or through the network connection. The storage drive can receive a program product, read programming steps recorded thereon, and transfer the programming steps into the memory for execution by the CPU. As noted above, the program product storage device can include any one of multiple removable media having recorded computer-readable instructions,

including magnetic floppy disks and CD-ROM storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor memory chips. In this way, the processing steps necessary for operation can be embodied on a program product.

Alternatively, the program steps can be received into the operating memory over the network. In the network method, the computer receives data including program steps into the memory through the network interface after network communication has been established over the network connection by well known methods that will be understood by those skilled in the art without further explanation.

The computer that implements the client side processing, and the computer that implements the server side processing, or any other computer device of the system, may comprise any conventional computer suitable for implementing the functionality described herein. FIGURE 9 is a block diagram of an exemplary computer device 900 such as might comprise any of the computing devices in the system. Each computer operates under control of a central processor unit (CPU) 902, such as an application specific integrated circuit (ASIC) from a number of vendors, or a "Pentium"-class microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. Commands and data can be input from a user control panel, remote control device, or a keyboard and mouse combination 904 and inputs and output can be viewed at a display 906. The display is typically a video monitor or flat panel display device.

The computer device 900 may comprise a personal computer or, in the case of a client machine, the computer device may comprise a Web appliance or other suitable Web-enabled device for viewing Web pages. In the case of a personal computer, the device 900 preferably includes a direct access storage device (DASD) 908, such as a fixed hard disk drive (HDD).

The memory 910 typically comprises volatile semiconductor random access memory (RAM). If the computer device 900 is a personal computer, it preferably includes a program product reader 912 that accepts a program product storage device 914, from which the program product reader can read data (and to which it can optionally write data). The program product reader can comprise, for example, a disk drive, and the program product storage device can comprise removable storage media such as a floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, a DVD disk, or the like. Semiconductor memory devices for data storage and corresponding readers may also be used. The computer device 900 can communicate with the other connected computers over a network 916 (such as the Internet) through a network interface 918 that enables communication over a connection 920 between the network and the computer device.

The CPU 902 operates under control of programming steps that are temporarily stored in the memory 910 of the computer 900. When the programming steps are executed, the pertinent system component performs its functions. Thus, the programming steps implement the functionality of the system illustrated in FIGURE 1. The programming steps can be received from the DASD 908, through the program product 914, or through the network connection 920, or can be incorporated into an ASIC as part of the production process for the computer device. If the computer device includes a storage drive 912, then it can receive a program product, read programming steps recorded thereon, and transfer the programming steps into the memory 910 for execution by the CPU 902. As noted above, the program product storage device can comprise any one of multiple removable media having recorded computer-readable instructions, including magnetic floppy disks, CD-ROM, and DVD storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor

memory chips. In this way, the processing steps necessary for operation in accord with the methods herein can be embodied on a program product.

Alternatively, the program steps can be received into the operating memory 910 over the network 916. In the network method, the computer receives data including program steps into the memory 910 through the network interface 918 after network communication has been established over the network connection 920 by well-known methods that will be understood by those skilled in the art without further explanation. The program steps are then executed by the CPU 902 to implement the processing of the system.

To implement the functionality described herein, it has been found that a suitable computer for performing database server tasks includes a "Pentium" level CPU having at least 128 MB of memory, 30 GB of disk storage, and 256 MB of disk swap space for files. A recommended configuration for computer performance would include, for example, a "Pentium III" processor at 700 MHz or faster, memory of 256 MB or greater, disk storage space of 50 GB or more, and swap space of 500 MB or more. A suitable configuration for performing user tasks as described above includes a "Pentium" level CPU having 128 MB memory, disk space of 240 MB with swap space of 256 MB, and an optional display circuit card supporting OpenGL and having 4 MB of memory. A recommended configuration includes, for example, a "Pentium III" processor at 500 MHz or faster, memory of 256 MB or greater, disk space of 500 MB or more, swap space of 500 MB or more, and an optional display card having 8 MB of memory or more, supporting resolution of 1024 x 768.

In a preferred embodiment, the software used in the computing system described above includes, for the server machine, operating system software such as "Windows NT Server 4.0" from Microsoft

Corporation, with Service Pack 5, Version 1280 (10 June 1999) or more recent, with database management server software such as, but are not limited to, "Oracle Server Standard Edition 8.1" from Oracle Corporation. The software used in a preferred embodiment of the user machine includes operating system software such as "Windows NT Workstation 4.0" from Microsoft Corporation, with Service Pack 5, version 1280 (10 June 1999) or more recent, as well as "Oracle Client Standard Edition Version 8.1" or higher. The client machine will also be compliant with the "Java" programming language (Java Runtime Environment 1.2.2). As will be known to those skilled in the art, other configurations may be suitable, depending on the applications being used and the computer performance desired.

**E. Computational phenotyping**

Also provided herein is a method designated computational phenotyping. Computational (also referred to herein as *in silico* phenotyping). This refers to the method in which a 3-D protein structure is generated from a given genotype and protein-drug binding analyses *in silico* (computationally) are performed in order to determine whether drug binding does (i.e. sensitive) or does not (i.e. resistant) take place. This type of analysis is contemplated to be performed for an individual patient or subject or groups thereof, such as ethnic groups, gender-based or age-based groups, particular species or groups thereof) to assess or select a drug for treatment of a particular disease or other such use, and is done to assess efficacy of a particular drug on a desired target, where the target exhibits polymorphisms. The following discussion and example, below, is with reference to HIV PR and RT, but it is understood that the methods and applications can be applied to any protein or gene product that exhibits polymorphic variation, and particularly to gene products that are drug targets.

Among the methods of computational phenotyping, there are three distinct methodologies that are clinically useful for determining either resistance or sensitivity to particular HIV-1 antiviral therapeutics. These are: genotyping, phenotyping, and *virtual* phenotyping. These methodologies are used to optimize the choice of therapeutics during the initiation of therapy, after drug failure, and/or during salvage therapy. Genotyping involves extracting the HIV viral RNA and amplifying all or part of the genes encoding the protease and reverse transcriptase proteins and sequencing them in order to assess the presence of resistance-associated mutations.

In phenotyping, the amplified sequences are instead sub-cloned into expression vectors and then tested for their replicative ability *in vitro* by transfecting them into cultured and/or established cell lines, such as, for example, human T cells, monocytes, macrophage, dendritic cells, Langerhans cells, hematopoeitic stem cells, HeLa, XC, Mm5MT, LTL, COS 7, NIH3T3, LTA, MCF-7, or other cells derived from human tissues and cells that which are the principal targets of viral infection in the presence or absence of antiviral drugs (see, *e.g.*, U.S. Patent No. 5,837,464; see, also EP 0852626; EP 1012334; and EP 0877937), *Virtual* phenotyping (ViroLogic, Inc.) is an interpretive service in which the phenotype of a specimen (i.e. of a plant, animal, pathogen, or human) is inferred from the specimen's genotype based upon an extensive correlative database of known genotypes and phenotypes. Such a correlative database must be updated constantly to maintain clinical accuracy.

Similar to *virtual* phenotyping, computational or in *silico* phenotyping infers phenotype based upon specimen genotype. Computational phenotyping is distinct from *virtual* phenotyping in that sensitivity or resistance to drugs is determined directly through protein-drug binding

analysis performed *in silico* and not through correlation with a database of known genotypes and phenotypes. The advantage of computational phenotyping is that new resistance conferring mutations can be discovered rapidly and in "real time" without the need for phenotyping to train the genotype. Moreover, *in silico* phenotypes are not subject to error caused from compensatory mutations which may act synergistically or anti-synergistically with resistance-associated mutations to increase, decrease, or reverse specific drug resistances. Computational phenotyping will generate information that can, for example, be presented in a report that is marketed within the *in vitro* diagnostics industry as an adjunct test/service to help optimize therapy and assist physicians, farmers, acadmenic institutions, government agencies, and industries with specimen treatment. Thus, a computer-based method for predicting clinical responses e.g. drug sensitivity or drug resistance in patients, plants, animals, pathogens, and microorganisms based on genetic polymorphisms is provided.

The genotypes used in the methods are obtained from any source, including, but are not limited to, from a plant, animal, pathogen, or mammal with the most preferred source being a mammal, paticularly a human for whom a particular drug treatment is contemplated, and is the genotype of the drug target, such as, as exemplified herein, HIV RT or PR from a particular infected individual. Other examplary drug targets are proteins, polypeptides, oligopeptides, including, but not limited to, a receptor, enzyme, hormone, and any such compound with which drugs or other ligands interact to bring about a biological response. For exemplification of this method, the protein considered is an enzyme, in particular HIV protease (PR) and reverse transcriptase (RT), which are therapeutic drug targets. Nucleic acid encoding the target from individual sample, such as blood sample or other body fluid sample from a

mammal, such as a human patient, is sequenced, and the 3-D structure thereof determined. The drug of interest is computationally tested to assess whether it interacts with the sample.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

## EXAMPLE 1

## BINDING CORRELATIONS OF MUTANT FORMS OF HCV PROTEASE WITH DIFFERENT INHIBITORS

This example provides the results of a theoretical study of NS3 protease complexes with two known peptide inhibitors (see SEQ ID Nos. 1 and 2; Ingallinella *et al.* ((1998) *Biochemistry 37*:8906-8914).

**Introduction**

During HCV replication, the final steps of processing are performed by a virially encoded chymotrypsin-like serine protease NS3. NS3 is an approximately 3000 amino acid protein that contains, from the amino terminus to the carboxy terminus, a nucleocapsid protein (C), envelope proteins (E1 and E2) and several non-structural proteins (NS1, 2, 3, 4a, 4b, 5a and 5b). NS3 is an approximately 68 kDa protein, encoded by approximately 1893 nucleotides of the HCV genome, and has two distinct domains: (a) a serine protease domain containing approximately 200 of the N-terminal amino acids; and (b) an RNA-dependent ATPase domain at the C-terminus of the protein. The NS3 protease is considered a member of the chymotrypsin family and is a serine protease that is responsible for proteolysis of the polypeptide (polyprotein) at the NS3/NS4a, NS4a/NS4b, NS4b/NS5a and NS5a/NS5b junctions responsible for generating four viral proteins during viral replication. This protease is inhibited by N-terminal cleavage products of substrate peptides. The NS3 protease, which is necessary for polypeptide processing and viral replication has been identified, cloned and expressed (see, e.g., U.S. Patent No. 5,712,145).

Active NS3 forms a heterodimer with a polypeptide cofactor NS4A. The crystal structure of NS3 with and without the NS4A cofactor is known (see, *e.g.*, Love *et al.* (1996) *Cell 87*:331-342; Habuka *et al.* (1997) *Jikken Igaku 15*:2308-2313; Yan *et al.* (1998) *Protein Sci.* 7:837-847, which provides the structure with NS4A).

The NS3 protease is a target for design of antiviral drugs. For example, a series of potent hexapeptide inhibitors of NS3 has been developed by optimization of the product inhibitors (Ingallinella *et al.* (1998) *Biochemistry 37*:8906-8914).

**Analyses**

Models of the complexes of NS3 with the two protease inhibitor peptides were obtained by flexible docking of the peptides into the active site of the crystal structure of NS3/4A, followed by evaluation of protein-peptide binding energies. The models were tested by *in situ* modification of the docked ligands. A qualitative agreement between the binding energies and inhibitor $IC_{50}$ values obtained from literature was found.

The peptides studied were:

| Sequence* | $IC^{50}$, nM | SEQ ID |
|---|---|---|
| Ac-$Asp^1$-D-$Glu^2$-$Leu^3$-$Ile^4$-$Cha^5$-$Cys^6$-COO- | 15 | 1 |
| Ac-$Asp^1$-L-$Glu^2$-$Leu^3$-$Ile^4$-$Cha^5$-$Cys^6$-COO- | 60 | 2 |

* Cha = β-cyclohexylalanine

In the modeling studies, it was assumed that:

the high-affinity inhibitory peptides 1 and 2 have a similar mode of binding to the active site of NS3;

the minimum binding pharmacophore includes the SH group of $Cys^6$ and carboxyl groups of $Asp^1$, $Glu^2$ and $Cys^6$; and

the side chains of residues 3, 4 and 5 may enhance binding by non-specific hydrophobic interaction with NS3.

**Methods**

**Initial structure of the NS3-peptide complex**

The crystal structure of NS3 with a peptide cofactor NS4A was obtained from the arts (Kim et al. (1996) Cell 87:343) and was used in the studies with peptide inhibitors. The crystal structure of NS3/NS4A was regularized using molecular mechanics described herein. Initial NS3-NS4-peptide complexes were constructed by placing the peptides into the NS3 binding site expected by structural homology to by other serine proteases:

the C-terminal carboxyl was placed near the oxyanion-stabilizing site (residues 137-139);

the side chain of $Cys^6$ was inserted into the hydrophobic cavity formed by L135, F154 and A157; and

the $\epsilon$-amino group of K136 was placed in contact with the C-terminal carboxyl (see, Kim et al. (1996) Cell 87:343, Steinkuhler *et al.* (1998) *Biochemistry* 37:8899).

**Monte Carlo simulations**

In order to optimize the complexes, Biased Based Probability Monte Carlo (BPMC) simulations (Abagyan et al. (1994) J. Mol. Biol. 235:983) were performed on the NS3-peptide complexes using the ICM program (commercially available from MolSoft, San Diego, CA) with ECEPP/3 force field and atomic solvation energies (Momany *et al.* (1975) J. Phys. Chem. 79:2361, Nemethy *et al.* (1992) J. Phys. Chem. 96:6472, Abagyan *et al.* (1997) Computer Simulations of Biomedical Systems: Theoretical and Experimental Applications, vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 363). The sampling method was BPMC with random change of one variable at a time. A Metropolis acceptance criterion was applied after energy minimization (quasi-Newton, up to 1000 steps). Simulations were performed at a temperature of 1000° K. The

peptide translational and rotational degrees of freedom, all peptide torsion angles and $\chi$ angles of the protein side-chains located within 7.0 Å of any peptide atom were varied during the BPMC simulations.

The energy function used in the MC simulations included:

ECEPP/3 terms for energy *in vacuo* (VDW (van der Waals), H-bond, electrostatic and torsion potentials);

distance dependent electrostatics with $e_0 = 4.0$; and

surface energy with atomic solvation parameters.

The total energies of the complexes were calculated including contributions from: ECEPP/3 VDW, H-bond, S-S bond and torsion terms; exact-boundary electrostatic energy with $e_0 = 8.0$; and side-chain entropies. Hydrophobic free energies were estimated as $sA$, where $A$ is accessible surface area and $s$ is a tension constant of 0.03 kcal/mol$Å^2$.

**Strategy of the flexible Monte Carlo docking**

The simulations proceeded with multiple, relatively short MC runs (2000-5000 generated structures). New docking cycles were started from the lowest-energy or other interesting structures found in previous runs. Structures saved during various MC runs were sorted by total energies and RMSD (root-mean-squared deviation), and compressed into a cumulative conformational stack. Binding energies were calculated for representative structures of each complex thus obtained. This strategy was more efficient than continuous long simulations because the variable torsion angles and distance constraints are defined for an initial structure and do not change during the MC run.

**Binding energies of the peptide-protein complexes**

For low-energy conformations found after several iterative BMPC cycles, peptide-protein binding energies were estimated using the equation:

$$E_{bind} = E_o + E_{compl} - E_{pept} - E_{prot},$$

where $E_{compl}$ is the energy of the complex, $E_{pept}$ & $E_{prot}$ are separate energies of the peptide and protein, respectively, and $E_o$ is an adjustable constant.

The binding energy function included: exact-boundary electrostatic free energy contributions; side-chain entropy; and surface tension hydrophobic free energy terms. (Zhou and Abagyan (1998) Folding Design 3:513, Schapira *et al.* (1999) J. Mol. Recognition 12:177). ECEPP/3 hydrogen-bonding terms were included with a weight of 0.5.

**Results**

**Models of the NS3-peptide complexes**

RMSD between pharmacophore atoms of peptides 1 and 2 were calculated for all pairs of BPMC structures. Two models of the NS3-peptide complexes were selected assuming (1) similar positions of pharmacophore groups of two peptides in the binding site (RMSD $\leq$ 2.0 Å) and (2) low binding energy of the complexes ($\Delta E_{bind}$ < 5.0 kcal/mol). Two models of the NS3-peptide complex were selected by visual inspection.

Characteristics of the binding sites for peptide inhibitors in two NS3-peptide complex models are summarized in **Table 1**.

**Table 1**

| site | Peptide residue | NS3 residue, group | Type of interaction | Present for Peptide Model 1 | Model 2 |
|---|---|---|---|---|---|
| P1 | $Cys^6COO^-$ | K136 $NH_3$+<br>G137 NH<br>S139 OH | H-bond/el.<br>H-bond<br>H-bond | 1,2<br>1,2<br>1,2 | 1,2<br>2<br>2 |
| | $Cys^6$ SH | L135, F154, A157 | hydroph | 1,2 | 1,2 |
| P2 | $Cha^5$ | H57, R155, A156<br>A157, V158 | hydroph<br>hydroph | 1,2<br>- | -<br>2 |
| P3 | $Ile^4$ | V132, S133<br>V158, C159 | hydroph<br>hydroph | 1,2<br>- | 2<br>1 |

| | | | | | |
|---|---|---|---|---|---|
| P4 | $Leu^3$ | Res. 157 to 160<br>V132, S133 | hydroph<br>hydroph | 1,2<br>- | 2<br>1 |
| P5 | $Glu^2$ COO- | R161 guanidine | H-bond/el. | - | 1,2 |
| P6 | $Asp^1$ COO- | R161 guanidine<br>S133 OH | H-bond/el.<br>H-bond | 1,2<br>- | -<br>1,2 |

**Validation of the models: modifications of the protein and ligands in the binding site**

In order to validate the proposed models, the K136M mutation and peptide modifications known from SAR (structure-activity relationship) studies were performed in low-energy structures of the NS3-peptide 2 complex.

Positions of the modified ligand and conformations of adjacent protein side chains were adjusted by energy minimization. Distance restraints were applied to keep the ligand near its initial position.

Changes in calculated binding energies upon modifications, $\Delta E_{bind}$ (calc), were compared to the values expected from ratios of inhibitory potencies, $\Delta E_{bind}(exp)$.

$$\Delta E_{bind}(exp) = RT\ ln(IC_{50}^{mod}/IC_{50}^{o}),$$

where $IC_{50}^{o}$ and $IC_{50}^{mod}$ are inhibitory potencies of the parent and modified compounds.

The correlation between experimental and calculated changes in binding energy upon ligand modifications in the binding site of NS3 is illustrated in

**FIG. 4.**

**Discussion**

The two NS3-peptide complex models suggest a common binding pattern for the inhibitor P1 site ($Cys^6$-OH) with the carboxyl group hydrogen-bonded to the oxyanion hole residues G137 and S139, and the

$Cys^6$ side chain embedded in a hydrophobic pocket formed by L135, F154 and A157.

This study confirms the possibility of hydrogen bonding between the C-terminal carboxyl and $\epsilon$-amino group of K136 suggested by Steinkuhler *et al.* ((1998) *Biochemistry* 37:8899) based on the K136M mutation in NS3. Changes in calculated binding energies upon mutation are consistent with an 8-fold increase in $K_1$ of an inhibitor with a free carboxyl group and with the lack of an effect on binding when the peptide is amidated.

The models differ in binding of the negatively charged side chains in positions P5 and P6. The R161 guanidine interacts with a carboxyl group of $Asp^1$ and $Glu^2$ in Models 1 and 2, respectively. In Model 2, the $Asp^1$ carboxyl also interacts with the hydroxyl of S133.

The models are in agreement with SAR data for peptide inhibitors of NS3. Predicted changes in binding energy upon modification of the protein and peptides correlate reasonably well with the changes expected from $IC^{50}$ ratios. Standard deviations of $\Delta E_{bind}$(calc) - $\Delta E_{bind}$(exp) were 0.8 and 1.6 kcal/mol for Models 1 and 2, respectively, with correlation coefficients of 0.62. After the largest outlier was removed from each dataset, correlations improved to 0.81 and 0.76, respectively.

**Conclusions**

An effective iterative Biased Probability Monte Carlo protocol for the docking of flexible peptide ligands into a flexible protein active site has been developed. Two models of the complexes of HCV NS3 protease with potent peptide inhibitors were proposed based on the docking simulations and on evaluation of protein-ligand binding energies. The models were validated by *in situ* modifications of NS3-peptide complexes and by correlation of binding energies of modified complexes with those expected from experimental $IC_{50}$ values. Proposed models can be used

for planning further mutagenesis studies of the HCV NS3 protease and the models can be used in the design of non-peptide inhibitors using structure-based drug design methodologies.

## EXAMPLE 2

## LEAD OPTIMIZATION BY RECEPTOR-BASED FREE ENERGY QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS (QSARS) FOR TNF RECEPTOR ANTAGONIST DISCOVERY

The goal of the modeling studies in this phase was to identify binding modes and complex structures of the compounds that bind to TNF receptor type I protein in order to guide the design of new compounds. An approach that relies on docking compounds to the receptor, evaluating free energy changes of binding of the docked structures, and comparing the calculated values with experimental inhibition constants $K_i$ of the compounds was developed. The success of the calculations was assessed by evaluating the consistency of the calculated free energy changes of binding and the experimental $K_i$.

The difference in free energy changes of binding between two compounds with inhibition constants $K_i$ and $K_i'$ can be calculated as,

$$\Delta\Delta G = -kT \ln K_i'/K_i$$

where k and T are Boltzmann's constant and absolute temperature, respectively.

The 13 active compounds were studied. Their potencies, as measured by $K_i$, range from 0.1 to 30 $\mu$M, spanning about 3 kcal/mol in free energy. It was found that the calculated free energy changes of binding are highly consistent with the corresponding experimental values, with correlation coefficient 0.966 and difference less than 0.5 kcal/mol (see Table 2 and Figure 4). The predicted binding modes and complex structures can thus be accepted with confidence.

To modify these compounds, important pharmacophore features on the surface of the receptor that are critical for binding of the compounds

were identified. These features include a hydrophobic belt, a hydrophilic belt and 3 hydrogen bond donor sites. A few of potential hydrogen bonding sites, which are not used by the current compounds, were also derived, and can be used for designing more potent binders.

Graphics-guided redesign of the compounds was performed. The free energy calculation was used to predict the binding activity of each design. Fourteen new compounds were thus designed and binding activities were predicted. The chemical structures of the designed molecules, together with the binding modes of the lead compounds, were synthesized and shown to have high affinity for the target. Some of them exhibit a $K_i$ in low-nanomolar range. Hence the method provided herein for modification of drugs for binding to calculated 3-D structures of a target protein resulted in redesigned drug candidates with enhanced affinity for the target.

This approach has advantages over the traditional x-ray crystallography method, which include the following:

(1) The binding modes are determined for a group of compounds instead of single compound; analysis of similarity and differences reveals rich information in binding mechanisms.

(2) The predictive power of the free energy calculation is very desirable for redesign of compounds.

(3) The correlation with the biochemical activities assures relevancy of the explored binding modes, while a structure given by x-ray crystallography may not necessarily be one related to the biological functions of the compound.

A comparison of calculated relative free energy changes of binding $\Delta\Delta A$ and experimental $\Delta\Delta G$ converted from inhibition constants $K_i$ (all in kcal/mol) of the compounds (referenced by a code name) is presented in **Table 2.**

Table 2

| Compound | ΔΔA | ΔΔG |
|---|---|---|
| SBI-2030 | 0 | 0 |
| SBI-2002 | -0.97 | -1.25 |
| SBI-2005 | -0.72 | -1.14 |
| SBI-307 | -0.56 | -0.08 |
| SBI-2008 | -0.53 | -0.88 |
| SBI-2006 | -0.34 | -0.44 |
| SBI-306 | -0.07 | 0.40 |
| SBI-2000 | 0.29 | 0.27 |
| SBI-2001 | 0.72 | 1.12 |
| SBI-304 | 1.55 | 1.45 |
| SBI-308 | 1.70 | 1.78 |
| SBI-305 | 1.86 | 1.67 |
| SBI-2048 | 1.95 | 1.94 |

A comparison of calculated *versus* experimental binding free energy changes is given in **FIG. 5.**

## EXAMPLE 3

**HIV Protease Models for Drug Studies**

Antiviral therapy for AIDS has focused on the discovery and design of inhibitors for two main enzyme targets of the HIV-1: reverse transcriptase (RT) and protease (PR). HIV RT is a heterodimer composed of p51 and p66 subunits. The p51 subunit is composed of the first 450 amino acids encoded by the RT gene and the p66 subunit is composed of all 560 amino acids of the RT gene. RT is responsible for RNA-dependent DNA polymerization, RNaseH activity, and DNA-dependent DNA polymerization.

HIV PR is a homodimer of two identical 99-amino acid chains. HIV PR is an aspartic proteinase that is responsible for the post-translational processing of the viral gag and gag-pol polyprotein gene products, which yields the structural proteins and enzymes of the viral particle (see, *e.g.*, Erickson *et al.*(1996) *Annu. Rev. Pharmacol. Toxicol. 36*:545-571, Bouras *et al.* (1999) *J. Med. Chem. 42*:957-962). Despite several promising new anti-HIV agents, the clinical emergence of drug-resistant variants of HIV limits the long-term effectiveness of these drugs. Genetic analysis of the resistant forms of HIV has identified a number of critical mutations in the RT and PR genes. Moreover, structural analysis of inhibitor-enzyme complexes and mutational modeling studies can lead to a better understanding of how these drug-resistant mutations exert their effects at the structural and functional levels.

**HIV-PR inhibitor computational binding studies**

This example provides the results of a computational study on HIV PR. The 3-D protease structure was generated, docked with known viral inhibitors, and analyzed via free energy of binding studies described herein. A quantitative agreement between the calculated add experimental protease-drug binding energies was obtained. Moreover, a series of 3-D HIV PR models were analyzed to identify the invariant regions of the protease. These insights have implications for the design of new drugs and therapeutic strategies to combat AIDS drug resistance.

**Optimization of 3D structures**

Five PR inhibitors approved by the FDA for clinical use were used: saquinavir, nelfinavir, indinavir, amprenavir, and ritonavir (Figure 6). Initial 3-D structures for the wild-type HIV PR complexes with these FDA approved inhibitors were obtained from the Protein Data Bank and were then optimized using Monte Carlo (MC) simulations with an ECEPP/3 force field as described in Example 1. The energy function used in the

MC simulations included: ECEPP/3 terms for energy in vacuo (van der Waals, H-bond, electrostatic and torsion potentials); distance dependent dielectrics with $e_0 = 4.0$; and surface free energy calculated using atomic solvation parameters ((Dudek *et al.* (1998) *J. Computational Chem.* *19*:548-573, Wang *et al.* (1995) J. Mol. Biol. 253:473-492). Standard ECEPP charges were used for the protein residues. Lys, Arg, Glu, and Asp residues were charged. Charged and protonated states of Asp 125 (chain B) were considered as well. The inhibitors were docked into the active site of the protease, and the protein-drug complexes were energetically refined using the methods described in Example 1. Partial charges for the inhibitors were calculated with the Gasteiger-Marsili method implemented in SYBYL 6.5 (Tripos Assoc., Inc.). Different protonation states were examined for indinavir and amprenavir, but the other inhibitors were assumed to be electroneutral. Water molecules located within 7.0 Å from a ligand atom in the X-ray structure were retained in the model complex during optimization.

**Calculation of binding energies**

For low energy conformations found after several iterative BMPC cycles, protein-drug binding energies were estimated using the equation:

$$E_{bind} = E_o + E_{compl} - E_{ligand} - E_{prot},$$

where $E_{compl}$ is the energy of the complex, $E_{ligand}$ & $E_{prot}$ are energies of the ligand and protein when separated, and $E_o$ is an adjustable constant. The binding energies of the protein and ligand were calculated using the following energy function:

$$E = E_{el} + E_{vw} + E_{hb} + E_s,$$

where $E_{el}$ is the exact-boundary electrostatic using $e_0 = 8.0$, $E_s$ is the side-chain entropy term, and $E_{vw}$ and $E_{hb}$ are the ECEPP/3 van der Waals and hydrogen-bonding terms.

After the energies of the wild type PR-inhibitor complexes were calculated, mutation sites were introduced into the optimized X-ray structures or model complexes. The amino acid substitutions were followed by local optimization, using an ECEPP/3 force field, of protein side chains around the mutation sites via the energy minimization of substructures that included the ligand, water molecules within the sphere of radius 7.0 Å around the ligand, and protease residues within the sphere of radius 3-5 Å around the mutated residues. The energy of binding of the mutated complex was calculated based on the equation described herein. The difference in binding energy resulting from mutations (mut) of the wild-type (WT) protease were calculated using the following equation:

$$\Delta E_{bind}(\text{calculated}) = E_{bind}(\text{WT}) - E_{bind}(\text{mut}).$$

This change in binding energy was compared to data from experimental (exptl) studies (Gulnik *et al.* (1995) Biochemistry 35:9282-9287, Klabe *et al.* (1998) Biochemistry 37:8735-8742, Pazhanisami *et al.* (1996) J. Biol. Chem. 271:17979-17985, Jacobsen *et al.* (1995) Virology 206:527-534, Maschera *et al.* (1996) J. Biol. Chem. 217:33231-33235) based on the equation:

$$\Delta E_{bind}(\text{exptl}) = RT\ln(K_i\text{mut}/K_i\text{wt}).$$

Plots of $\Delta E_{bind}$(calculated) vs. $\Delta E_{bind}$(exptl) were generated, and the results, summarized in Table 3, show a strong correlation between the calculated binding energies and the experimentally determined binding energies for the PR-inhibitor complexes. For example, the correlation coefficient R for PR-ritonavir and PR-amprenavir is 0.9, where R = 1 denotes congruency between the computationally calculated and experimentally determined binding energy data. These correlation data validate the computational protocol and calculations described herein as a method for predicting protein-drug binding or protein-drug resistance (i.e. non-binding). The

evaluation of changes in binding energy of protein-drug complexes upon protein sequence variations can be used as a possible descriptor and, thus, can be used to predict the efficacy of drugs on proteins resulting polymorphisms in genes. Moreover, the analysis of the free energy of binding in complexes between the protein models that are produced by the method set forth in this example and drugs that have been designed or modified is a good predictive tool for drug designers.

**TABLE 3**

Correlation between Experimental and Calculated Binding Energies for HIV Protease Inhibitors

| HIV PRInhibitor | X-ray Complex ID | No of exptl. data points | Correlation coefficient R | Correlation S.D., kcal/mol |
|---|---|---|---|---|
| Saquinavir | 1HXB | 18 | 0.84 | 0.68 |
| Indinavir | 1HSG | 17 | 0.79 | 0.80 |
| Ritonavir | 1HXW | 12 | 0.90 | 0.72 |
| Amprenavir | 1HPV | 15 | 0.90 | 0.54 |
| Nelfinavir | 1OHR | Insufficient data | | |

**Identification of structural invariant regions of HIV Protease**

Clinical effectiveness of HIV PR inhibitors is limited by the rapid emergence of drug-resistant mutations. Resistant PR variants first occur by the mutation of amino acids close to or in and around the drug binding site, which are then accompanied by compensatory mutations of more distant amino acids. The identification of highly conserved, structural invariant regions of a PR would provide new potential targets and thus lead to the development of therapeutics having greater clinical efficacy than those drugs commonly employed to treat HIV.

The protein sequences of HIV protease were obtained from GenBank and from the blood samples of patients using standard isolation and sequencing techniques well known in the arts. The protein sequences were modeled into 3-D structures using the computational

protocol described in Example 1. The protease sequences were aligned, and the frequency of mutation, regardless of type, was determined at each amino acid position and plotted in Figure 7, where the frequency of mutation in this set of HIV-1 Protease sequences varied from 0 to 40%. Sequence alignment also revealed how many different types of amino acids could be substituted in any specific residue, yielding the tolerance of each residue to substitutions of different types. The data showing the frequency of mutation of each residue out of PR sequences, the types of mutations, and the distance of the mutating residue from the active site (Asp 28) are shown in FIG. 8. This information, sequences obtained from 10591 different genotypes, was used to identify invariant and/or highly conserved regions of PR and to map these regions to a 3-D structure for the purpose of identifying new potential regions on the protein as targets for therapeutic intervention. These invariant regions include, but are not limited to, residues 1-9, 25-29, 49-52, 78-81, and 94-99, where residue 1 is an aliphatic amino acid, more preferably proline; residue 2 is a hydrophilic amino acid, more preferably glutamine; residue 3 is an aliphatic amino acid, more preferably isoleucine; residue 4 is a hydrophilic amino acid, more preferably threonine; residue 5 is a hydrophobic amino acid, more preferably leucine; residue 6 is an aromatic amino acid, more preferably tryptophan; residue 7 is a hydrophilic amino acid, more preferably glutamine; residue 8 basic amino acid, more preferably arginine; residue 9 is an aliphatic amino acid, more preferably proline; residue 25 is a hydrophilic amino acid, more preferably aspartic acid; residue 26 is a hydrophilic amino acid, more preferably threonine; residue 27 is an aliphatic amino acid, more preferably glycine; residue 28 is an aliphatic amino acid, more preferably alanine; residue 29 is an acidic amino acid, more preferably aspartic acid; residue 49 is an aliphatic amino acid, more preferably glycine; residue 50 is a hydrophobic amino acid,

more preferably isoleucine; residue 51 is an aliphatic amino acid, more preferably glycine; residue 52 is an aliphatic amino acid, more preferably glycine; residue 78 is an aliphatic amino acid, more preferably glycine; residue 79 is an aliphatic amino acid, more preferably proline; residue 80 is a hydrophilic amino acid, more preferably threonine; residue 81 is an aliphatic amino acid, more preferably proline; residue 94 is an aliphatic amino acid, more preferably glycine; residue 95 is a thio-containing amino acid, more preferably cysteine; residue 96 is hydrophilic amino acid, more preferably threonine; residue 97 is hydrophobic amino acid, more preferably leucine; residue 98 is hydrophilic amino acid, more preferably asparagine; and residue 99 is an aromatic amino acid, more preferably phenylalanine. These invariant regions can subsequently be used to assist in the design drugs or therapeutic agents which bind to the invariant regions and disrupt the activity of the protease with greater efficacy than drugs commonly used to treat HIV and where the free energy of binding between said drug or therapeutic agent and the structural invariant region is evaluated as described herein. The methods described in this example can also be applied to HIV RT and to any protein of interest that exhibits polymorphisms.

## EXAMPLE 4

**Computational Phenotyping of HIV-1 Protease and Reverse Transcriptase**

Computational or *in silico* phenotyping is performed to assess phenotypic properties of a protein. This example demosntrates application of this method to HIV-1 protease and reverse transcriptase to test whether the efficacy of various protease inhibitors for an HIV patient.

To practice this method 3-D structures of HIV-1 protease and reverse transcriptase based upon the nucleic acid isolated from HIV from a patient are generated. Protein-drug binding analysis *in silico* in order to

determine whether drug binding does (i.e. sensitivity) or does not (i.e. resistance) take place.

Sequencing of HIV-1 Protease and Reverse Transcriptase is performed on HIV-1 cDNA following extraction, reverse transcription, and PCR amplification of viral RNA obtained from patient specimens, such as blood samples or other body fluid or tissue samples. Methods for the extraction, reverse transcription, and PCR amplification of viral RNA are well known in the art. For each sequence, a computer-generated 3-D structure of the protein is modeled and then docked with antiviral drugs in silico using methods described in Example 1 and elsewhere herein to analyze protein-drug interactions. Antiviral drugs that can be tested include, but are not limited to, saquinavir, indinavir, ritonavir, amprenavir, and nelfinavir for HIV protease; zidovudine, lamivudine, stavudine, zalcitabine, didanosine, abacavir, adefovir, delavirdine, nevirapine, and efavirenz for HIV reverse transcriptase; and any FDA-approved or non-FDA approved antiviral drug. From these protein-drug interaction studies, relative drug resistance or sensitivity is inferred by calculating and evaluating the free energy of binding in low energy conformations of complexes between the variant protease structure and docked antiviral drug or variant reverse transcriptase structure and docked antiviral drug, using the methods described in Examples 1 and 3 and elsewhere herein.

The results of the computational phenotyping procedure can be presented as a patient report that states whether a drug or drugs are sensitive or resistant to the RT or PR obtained from the patient. Such a patient report assists physicians in selecting appropriate drugs for HIV patients. It also is useful for the *in vitro* diagnostics industry in an adjunct test/service capacity to help optimize antiviral therapy.

## EXAMPLE 5

**HIV Protease and Reverse Transcriptase Databases**

Exemplary databases of the 3-D protein structures of polymorphic variants are described in this example. The HIV PR and RT databases are a comprehensive collection of 3-D polymorphic structural data along with related information, including nucleic acids encoding all or a portion of the protein. These data provide a means to understand differences in the interactions between a drug or drugs and the structural variations of the drug targets.

This example describes the creation, interface for, and use of structural variant databases of HIV protease and reverse transcriptase polymorphic variants.

**Construction of databases**

To implement the RT or HIV database described herein, suitable computer for performing database server tasks includes a "Pentium" level CPU having at least 128 MB of memory, 30 GB of disk storage, and 256 MB of disk swap space for files. A recommended configuration for better computer performance would include, for example, a "Pentium III" processor at 700 MHz or faster, memory of 256 MB or greater, disk storage space of 50 GB or more, and swap space of 500 MB or more. A suitable configuration for performing user tasks as described above includes a "Pentium" level CPU having 128 MB memory, disk space of 240 MB with swap space of 256 MB, and an optional display circuit card supporting OpenGL and having 4 MB of memory. A recommended configuration for better performance would include, for example, a "Pentium III" processor at 500 MHz or faster, memory of 256 MB or greater, disk space of 500 MB or more, swap space of 500 MB or more, and an optional display card having 8 MB of memory or more, supporting resolution of 1024 x 768.

Preferably, the software used in the computing system described above includes, for the server machine, operating system software such as "Windows NT Server 4.0" from Microsoft Corporation, with Service Pack 5, Version 1280 (10 June 1999) or more recent, with database management server software such as "Oracle Server Standard Edition 8.1" from Oracle Corporation, or better. The software used in a preferred embodiment of the user machine includes operating system software such as "Windows NT Workstation 4.0" from Microsoft Corporation, with Service Pack 5, version 1280 (10 June 1999) or more recent, as well as "Oracle Client Standard Edition Version 8.1" or better. The client machine will also be compliant with the "Java" programming language (Java Runtime Environment 1.2.2). As will be known to those skilled in the art, other configurations may be suitable, depending on the applications being used and the computer performance desired.

**Database Interface**

The database interface was a Java-based interface with useful features. The database is interfaced to a molecular graphics package that includes 3-D visualization, including wire-frame representations; secondary structure ribbons; and solid surfaces, and structure analysis tools. The database also provides an interface to access all of the collected files from the same 3-D structure. The database interface also provides access to other databases, such as databases of chemical structures and public domain databases such as GenBank and the Protein Data Bank. The OpenGL and C++ module has real-time interaction with the sequence display and sequence analysis modules, such that highlighting residues in one display results in highlighting those same residues in other displays.

The relational database containing the protein information may be structured according to relational objects to facilitate the analysis and

computation processes described in the preceding examples. FIG. 10 is a graphical representation of the database objects for the system described herein. The database is organized by classes, each of which is characterized by data attributes and subclasses for the proteins.

FIG. 10 shows that the database design includes classes comprising Variant and related classes of Sample, Residue, Model, Resistance_Entry, and Protein. Other classes include Conformation, Residue_Conformation, Atom, Drug, Family, and Subfamily. These classes store attribute data values and specify class parameters and behaviors to provide the functionality described herein.

For example, FIG. 10 shows that the Variant class stores parameters to specify a variant, including subclasses that specify a Variant_ID, Sample_ID, Protein_ID, Name, and Sequence, where Variant_ID is the identification number of the variant; Sample_ID is the identification number of the sample from which HIV PR and RT were obtained; Protein_ID is the identification number of the protein i.e. PR or RT; Name is the name of the variant distinguishing it from other variants encoded by the same DNA due to ambiguities in the nucleic acid sequence; and Sequence is the nucleotide or amino acid sequence. Similarly, FIG. 10 shows that the Sample class includes subclasses relating to a specific sample and which specify Sample_ID, Sample_Date, Sex, Ambiguity_Number, Distance, Sequence_Length, Sequence, Clade, and Region, where Sample_ID is as defined herein; Sample_Date is the date the sample was obtained; Sex is the gender of the sample donor; Ambiguity_Number is fraction of ambiguous nucleotide positions; Distance is a normalized number the variation of an amino acid from the master clade; Sequence_Length is the length of the sequence; Sequence is as defined herein; Clade is the master sequence; and Region is the geographic location from which the sample was obtained. The Model

class includes subclasses comprising Model_ID, Model_Name, Variant_ID, and Drug_ID, where Model_ID is the identification number of the 3-D protein model; Model_Name is the name of the 3-D protein model; Variant_ID is as defined herein; and Drug_ID is the identification number of the drug i.e. antiviral drug. The atom class includes the subclasses comprising Atom_Name, Residue_Conformation_ID, X_Coordinate, Y_Coordinate, and Z_Coordinate, where Atom_Name is the name of atom in the 3-D protein structure; Residue_Conformation_ID is the identification number of the amino acid conformation in a 3-D structure; and X_Coordinate, Y_Coordinate, and Z_Coordinate are the coordinates of the 3-D protein structure. The conformation class includes the subclasses comprising Conformation_ID, Model_ID, and Refinement_Level, where Conformation_ID is the identification number of a conformation of a 3-D structure; Model_ID is as defined herein, and Refinement_Level is the number of times the conformation was refined energetically. The drug class includes the subclasses comprising Drug_ID, Profile, Symbol, Name1, Name2, Company, and URL, where Drug_ID is as defined herein; Symbol is the FDA symbol for the drug; Name1 is the name of the drug, Name2 is an alternative name of the drug; Company is the company that makes the drug; and URL is the website address of the company that makes the drug. The residue_conformation class includes the subclasses comprising Residue_Conformation_ID, Conformation_ID, and Residue_ID, where Residue_Conformation_ID is as defined herein; Conformation_ID is as defined herein; and Residue_ID is the identification number of the amino acid. The Resistance_Entry class includes the subclasses comprising Resistance_Entry_ID, Profile, Protein_ID, Residual_Number, Amino_Acid, Weight, and Maximum_Weight, where Resistance_Entry_ID is; Protein_ID is as defined herein, Amino_Acid is the amino acid. The Family class includes the subclasses comprising Family_ID and

Family_Name, where Family_ID is the identification number of the protein family and Family_Name is the name of the protein family. The SubFamily class includes the subclasses comprising SubFamily_ID, SubFamily_Name, and Family_ID, where SubFamily_ID is the identification number of the protein subfamily, SubFamily_Name is the name of the protein subfamily, and Family_ID is as defined herein. The Protein class includes the subclasses comprising Protein_ID, Protein_Name, Species, Multiple_Domain, Multiple_Chain, and Wild_Type, where Protein_ID is as defined herein, Protein_Name is the name of the protein i.e. RT or PR; Species is the species of the source of the protein i.e. humans; Multiple_Domain is the domain of the protein i.e p66 or p51 in the case of RT; Multiple_Chain is the a or b chain in the dimers of RT and PR; and Wild_Type is the wild-type protein sequence for RT and PR. The residue class includes the subclasses comprising Residue_ID, Variant_ID, Chain, Residue_Number, Insertion_Code, and Residue_Code, where Residue_ID is the identification number of the amino acid, Variant_ID is as defined herein, Chain, Residue_Number is the numbering of an amino acid in a protein sequence, Insertion_Code is the identification number if different insertions occur in the amino acid sequence, and Residue_Code is the single letter or 3-letter code of an amino acid. Those skilled in the art will understand the database design exemplified in FIG. 10. It should be understood that other classes or parameters may be included, as selected by those skilled in the art, for the desired database design.

**Database Content**

The databases contain information on the variants of HIV PR and RT present in patient populations. The master amino acid sequence, nucleic acid sequence, and 3-D structure are obtained from GenBank; an exemplary master sequence is set forth in SEQ ID No. 118. Nucleotide sequences exhibiting polymorphisms and the corresponding structural

variant protein sequences are determined by isolating nucleic from viruses and viral nucleic acid obtained from the blood samples of patients throughout the US, as well as from other countries, using sequencing methods well known in the art. The sequences were inputted into the RT and PR databases. Exemplary of the nucleotide sequences and the encoded amino acids for HIV RT and PR in this data base are set forth in SEQ ID NOS. 3 to 117, where r is g or a; y is t/u or c; m is a or c; k is g or t/u; s is g or c; w is a or t/u; b is g or c or t/u; d is a or g or t/u; h is a or c or t/u; v is a or g or c; and n is a or g or c or t/u or unknown or other. The amino acid sequences of the wild type and structural variants are used to create 3-D protein structures which are deposited into the databases.

**1. 3-D Protein Models**

The structure of the wild-type or master sequence model of PR and RT were obtained from the crystal structures found in PDB. The initial structure was refined energetically using BPMC with an ECEPP force field as described in Example 1. The quality of the model was assessed by calculating Normalized Residue Energies (NREs), where models with $e_{av} \geq 1.5$ require further energetic refinement; and models with $e_{av} < 1.5$ were deposited into the database as described herein. The 3-D protein structures of the variant sequences were generated by comparing these structures to the master sequence (see, *e.g.*, SEQ ID No. 118; *i.e.*, homology modeling) and energetically refining the models *ab initio*, using the same force field and BPMC procedure as the master sequence and applying the same quality control standard as described herein. Figure 11 is a tabulation of the 3-D coordinates of an exemplary HIV PR entry in a database that includes 3-D structures. For US purposes and where permitted, Tables 4 and 5 are provided electronically on CD ROM. These Tables house the coordinates that represent the 3-D protein structures of

proteins encoded by the nucleic acids set forth in SEQ. ID. NOS. 3-117. It will be noted that these sequences encode a full length PR and about 200 nucleotides the p51 subunit, which is the subunit of interest herein. To construct the full-length 3-D structure, the 3-D structure of each encoded portion of the p51 subunit was generated and then combined with the structure of the master sequence to produce a full-length structure.

These 3-D structures in the database can be selected and exported into computational docking programs for analyzing protein-drug interactions on known drugs, new drugs or modified drugs. The database can be mined to find protein models that correspond to patients with a particular genetic polymorphism, patients with the most commonly occurring polymorphism, to a relevant patient subpopulation (*e.g.*, gender, age, race, or other characteristic), to patients receiving a specific treatment regimen, to patients exhibiting a particular clinical response, to structural invariants, or to other relevant criteria.

Drugs can be docked into the active sites of PR and RT and subsequently energetically refined using an ECEPP force field and BPMC as described in Example 1. The quality control is that the protein-drug complex represents a low energy conformation, which may take several iterative BMPC cycles. Then, the binding energies of the protein-drug complexes can be estimated using the methods of Example 1. Drug designers can modify the structures of drugs

or design new drugs, using methods well known in the arts, to maximize the drug binding to the models generated by this database.

**2. Other Data**

Each PR or RT nucleotide sequence in the database has associated with it an identification number, the nucleotide sequence length, the translated amino acid sequence (or sequences in cases of ambiguous

nucleotide positions), a 3-D structure for each amino acid sequence (from which a number of structurally related values are calculated), the genotyping date, the gender of the patient, the geographical location from which the sample was sent, the clade of the sequence, the fraction of ambiguous nucleotide positions, drug information, and other clinical information.

**Database Usage**

A query menu allows the user to retrieve data based on the various fields: sample ID, residue number (with or without specific amino acid mutation), date gender, geographic location, distance from the master sequence, and other useful queries. The set of sequences that satisfies the user's query are brought up in a sequence display module, which have variations from the master sequence indicated initially, although the sequences can be highlighted according to predicted resistance. This subset of sequences can be subjected to further analyses. For example, a histogram summarizing the number of mutations at each position in the subset can be generated. The 3-D structures for any of the variants in the database can be displayed and analyzed in the structure visualization module, allowing the user to compare the similarities and differences between 3-D structures by superimposing the 3-D structures. The user and also export these structures into programs for protein-binding studies as described herein. Thus, by mining the databases, a user will access 3-D structures and clinical and sample information that can be used in and correlated with protein-drug binding studies of HIV PR and RT.

**Database Applications**

The HIV PR and RT databases have many applications. The applications include, but are not limited to, any application and method provided herein, such as databases that assist in de novo drug design and drug binding calculations. In particular, the database can be used in the

design of 2nd and 3rd generation drugs to combat potential resistance to HIV therapy, and it can be used in the design of drugs that will impact a broad spectrum of the infected population. The databases provide the ability to design drugs that focus on the most highly conserved regions of a drug target and drugs that will avoid resistance to mutation. The database could be used to rank drug candidates by likely efficacy within a given subpopulation of patients (e.g. age, race, gender) in pre-clinical trials and to predict the most effective drug regimen to give a patient, and for designing clinical trials.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

**CLAIMS**

1. A computer-based method of drug design based on genetic polymorphisms, comprising:

obtaining more than one amino acid sequence of target proteins that are the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-dimensional (3-D) protein structural variant models from the sequences; and

based upon the structures of the 3-D models, designing drug candidates, modifying existing drugs, identifying potential drug candidates or identifying modifications of existing drugs based on predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants.

2. The method of claim 1, wherein the structure-based drug design method comprises:

computationally docking the drug candidate or modified drug molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug candidate or modified drug molecules and the structural variants; and

designing and identifying drugs or modifications to existing drugs based on the binding interactions.

3. The method of claim 2 wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

4. The method of claim 1 wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models, wherein

the conserved structural features are used as a basis for structure-based drug design studies.

5. The method of claim 4, wherein the conserved structural features are stretches of non-contiguous residues, wherein each stretch contains at least two amino acids.

6. The method of claim 5, wherein the protein is human immunodeficiency virus protease.

7. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:

residue 1 is an aliphatic amino acid; residue 2 is a hydrophilic amino acid; residue 3 is an aliphatic amino acid; residue 4 is a hydrophilic amino acid; residue 5 is a hydrophobic amino acid; residue 6 is an aromatic amino acid; residue 7 is a hydrophilic amino acid; residue 8 is a basic amino acid; residue 9 is an aliphatic amino acid; residue 25 is an acidic amino acid; residue 26 is a hydrophobic amino acid; residue 27 is an aliphatic amino acid; residue 28 is an aliphatic amino acid; residue 29 is an acidic amino acid; residue 49 is an aliphatic amino acid; residue 50 is a hydrophobic amino acid; residue 51 is an aliphatic amino acid; residue 52 is an aliphatic amino acid; residue 78 is an aliphatic amino acid; residue 79 is an aliphatic amino acid; residue 80 is a hydrophilic amino acid; residue 81 is an aliphatic amino acid; residue 94 is an aliphatic amino acid; residue 95 is a thio-containing amino acid; residue 96 is a hydrophilic amino acid; residue 97 is hydrophobic amino acid; residue 98 is hydrophilic amino acid; and residue 99 is an aromatic amino acid.

8. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:

residue 1 is proline; residue 2 is glutamine; residue 3 is isoleucine; residue 4 is threonine; residue 5 is leucine; residue 6 is tryptophan; residue 7 is glutamine; residue 8 is arginine; residue 9 is proline; residue 25 is aspartic acid; residue 26 is threonine; residue 27 is glycine; residue 28 is alanine; residue 29 is aspartic acid; residue 49 is glycine; residue 50 is isoleucine; residue 51 is glycine; residue 52 is glycine; residue 78 is glycine; residue 79 is proline; residue 80 is threonine; residue 81 is proline; residue 94 is glycine; residue 95 is cysteine; residue 96 is threonine; residue 97 is leucine; residue 98 is asparagine; and residue 99 is phenylalanine.

9. The method of claim 6, wherein the HIV protease has the sequence of amino acids set forth in any of SEQ ID Nos. 3-74 and 77-117.

10. The method of claim 9, wherein the residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99.

10. The method of claim 1, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.

11 The method of claim 1, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a selected patient subpopulation.

12. The method of claim 1 wherein the structural variant models are stored in a relational database, comprising:

3-D molecular coordinates for the structural variants;

a molecular graphics interface for 3-D molecular structure visualization; computer functionality for protein sequence and structural analyses; and

database searching tools.

13. The method of claim 12, wherein the database further comprises one or more of observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.

14. The method of claim 1, wherein:

after generating the 3-D protein structural variant models, the method comprises:

computationally docking drug molecules with the target protein models; and

energetically refining the docked complexes; and

wherein the candidate drugs are specific for a protein with a selected polymorphism or specifically interact with all proteins exhibiting a polymorphism.

15. The method of claim 14, wherein the structure-based drug design method comprises:

computationally docking drug or potential new drug candidate molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug or potential new drug candidate molecules and the structural variants; and

designing potential new drugs or modifications to existing drugs based on the binding interactions.

16. The method of claim 15, wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

17. The method of claim 14, wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models; and

the conserved structural features are used as a basis for structure-based drug design studies.

18. The method of claim 17, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.

19. The method of claim 17, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a specific patient subpopulation.

20. The method of claim 12, wherein the selected model structures represent structural variants derived from patients the receive a specific treatment regimen.

21. The method of claim 12, wherein the selected model structures represent structural variants derived from patients that exhibit a particular clinical responses to a given drug.

22. The method of claim 12, wherein the selected model structures represent structural variants derived based on the duration of a particular drug treatment.

23. The method of claim 12, wherein the structural variant models are stored in a relational database, comprising:

3-D molecular coordinates for the structural variants;

a molecular graphics interface for 3-D molecular structure visualization; and

functionality for protein sequence and structural analysis; and

database searching tools.

24. The method of claim 12, wherein the database further comprises observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.

25. A computer-based method of selecting drug therapies for patients based on genetic polymorphisms, comprising:

obtaining amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

computationally docking drug molecules with the target protein models;

energetically refining the docked complexes;

determining the binding interactions between the drug or potential new drug candidate molecules and the models; and

selecting drug therapies based on the drug or drugs that have the most favorable binding interactions with the structural variant models.

26. The method of claim 25, wherein the binding interactions are determined by:

calculating the free energy of binding between the protein structural variant and the docked drug molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

27. The method of claim 1, further after generating the 3-D structural variant models, exporting some or all of them models into a program that computationally docks the models with test compounds to assess intermolecular interactions.

28. A computer-based method for predicting clinical responses in patients based on genetic polymorphisms, comprising:

obtaining one or more amino acid sequences for a target protein that is the product of a gene exhibiting genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

building a relational database of protein structural variants derived based on genetic polymorphisms and observed clinical data associated with particular polymorphisms exhibited in the patients, wherein the database comprises:

3-D molecular coordinates for the structural variant models;

a molecular graphics interface for 3-D molecular structure visualization;

computer functionality for protein sequence and structural analysis;

database searching tools; and

observed clinical data associated with the genetic polymorphisms, subject medical history and subject history associated with the genetic polymorphisms;

obtaining a target protein structural variant based on the same gene associated with a polymorphism in a patient;

generating a 3-D protein model based on the subject's gene sequence;

screening/comparing the 3-D model derived from the subject to the structures contained in the database by:

identifying structures in the database that are similar to the model derived from the subject; and

predicting a clinical outcome for the patient based on the clinical data associated with the identified structures.

29. A computer-based method for designing therapeutic agents that are active against biological targets that have become drug resistant due to genetic mutations, comprising:

obtaining a first 3-D protein structural variant model of a target protein against which a given drug has biological activity;

generating a second 3-D protein structural variant model of the target in which genetic mutations have occurred and against which the same drug is no longer biologically active;

comparing the structures of the first and second model to identify structural differences; and

performing structure-based drug design calculations in order to identify new drugs or modifications to the existing drug to bring about biological activity against the second model.

30. A computer-based method for identifying compensatory mutations in a target protein, comprising:

obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, wherein the structure of a form of the target protein that responds to a particular drug, including the active site, has been structurally characterized;

generating a 3-D structural model of the mutated protein;

comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations;

comparing the biological activities of the drug against both the mutated protein and the form of the protein that responds to the drug to determine the effects of the mutations on drug response; and

identifying the mutations in the protein that affect biological activity based on the comparisons.

31. A method for creating a 3-D structural polymorphism relational database, comprising:

obtaining one or more amino acid sequences of a target protein that is the product of a gene exhibiting a genetic polymorphism, wherein sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

energetically refining the models;

evaluating the quality of the models;

optionally obtaining associated clinical properties or data; and

inputting the model and any associated properties and/or data into a relational database.

32. The method of claim 31, wherein after energetically refining the models, the models are further refined.

33. The method of claim 31, wherein the database comprises amino sequences of two or more polymorphic variants.

34. The method of claim 31, wherein the database comprises amino sequences of ten or more polymorphic variants .

35. The method of claim 31, wherein the database comprises amino sequences of about 100 or more polymorphic variants .

36. The method of claim 31, wherein the database comprises amino sequences of about 1000 or more polymorphic variants .

37. The method of claim 31, wherein the database comprises amino sequences of more than 8000 polymorphic variants.

38. A database created by the method of claim 31.

39. The database of claim 38, comprising variant 3-dimensional structures of a selected target.

40. The database of claim 38 that comprises structures of proteases or polymerases.

41. The database of claim 38, wherein the proteases are viral proteases or polymerases.

42. The database of claim 38, wherein the viral proteases are human immunodeficiency virus proteases and the polymerase is a viral reverse transcriptase.

43. The method of claim 31, wherein quality is assessed by computing the normalized residue energies such that if $e_{av}$ is $\geq 1.5$ a model is further refined until $e_{av}$ is $< 1.5$; if $e_{av}$ is $< 1.5$ a model is deposited into the database.

44. The method of claim 1, wherein the target is an enzyme.

45. The method of claim 44, wherein the enzyme is a protease or polymerase.

46. The method of claim 45, wherein the polymerase is a reverse transcriptase.

47. The method of claim 44, wherein the target is a protein expressed by an infectious agent.

48. The method of claim 44, wherein the target is enzyme expressed by a an infectious agent.

49. The method of claim 48, wherein the agent is a human immunodeficiency virus (HIV).

50. A computer system, comprising a database containing data representative of the three dimensional structure of polymorphic variants of a drug target.

51. The system of claim 50, wherein the target is a cell surface receptor or an enzyme.

52. The system of claim 50, wherein the enzyme is a protease or a polymerase.

53. A database, comprising:

sequences of nucleotides encoding a protein or portions thereof, wherein proteins comprise polymorphic variants; and the portions encode a domain of the protein that comprises a site in the protein that binds to a drug candidates; and

the coordinates of 3-dimensional (3-D) structures of the encoded proteins or portions thereof.

54. The database of claim 53 that is a relational database.

55. The database of claim 53 that comprises at least 2 polymorphic variants and the corresponding 3-D structures.

56. The database of claim 55 that comprises at more than 10, more than 100, more than 1000, more than 8000, or more than 10,000 polymorphic variants and the corresponding 3-D structures.

57. The database of claim 53, wherein the protein is a receptor or enzyme from a eukaryotic or prokaryotic organism.

58. The database of claim 53, wherein the organism is a pathogen or a mammal.

59. The database of claim 53, wherein the organism is a pathogen is a virus or bacterium and the mammal is a human.

60. The database of claim 53, wherein the protein is a protease or a reverse transcriptase.

61. A database, comprising the sequences of nucleotides set forth in SEQ ID Nos. 3-117 that encode HIV protease or the portion of HIV reverse transcriptase set forth in each SEQ ID.

62. The database of claim 53, further comprising 3-D structural coordinates for a protein or portion thereof comprising sequences of amino acids encoded by each of SEQ ID Nos. 3-117.

63. The database of claim 54, wherein the protein is HIV protease.

64. The database of claim 54, wherein the protein is HIV reverse transcriptase.

65. The method of claim 1, wherein the target protein is a eukaryotic or prokaryotic protein.

66. The method of claim 1, wherein the target protein is an animal protein, a plant protein or a protein from a pathogen.

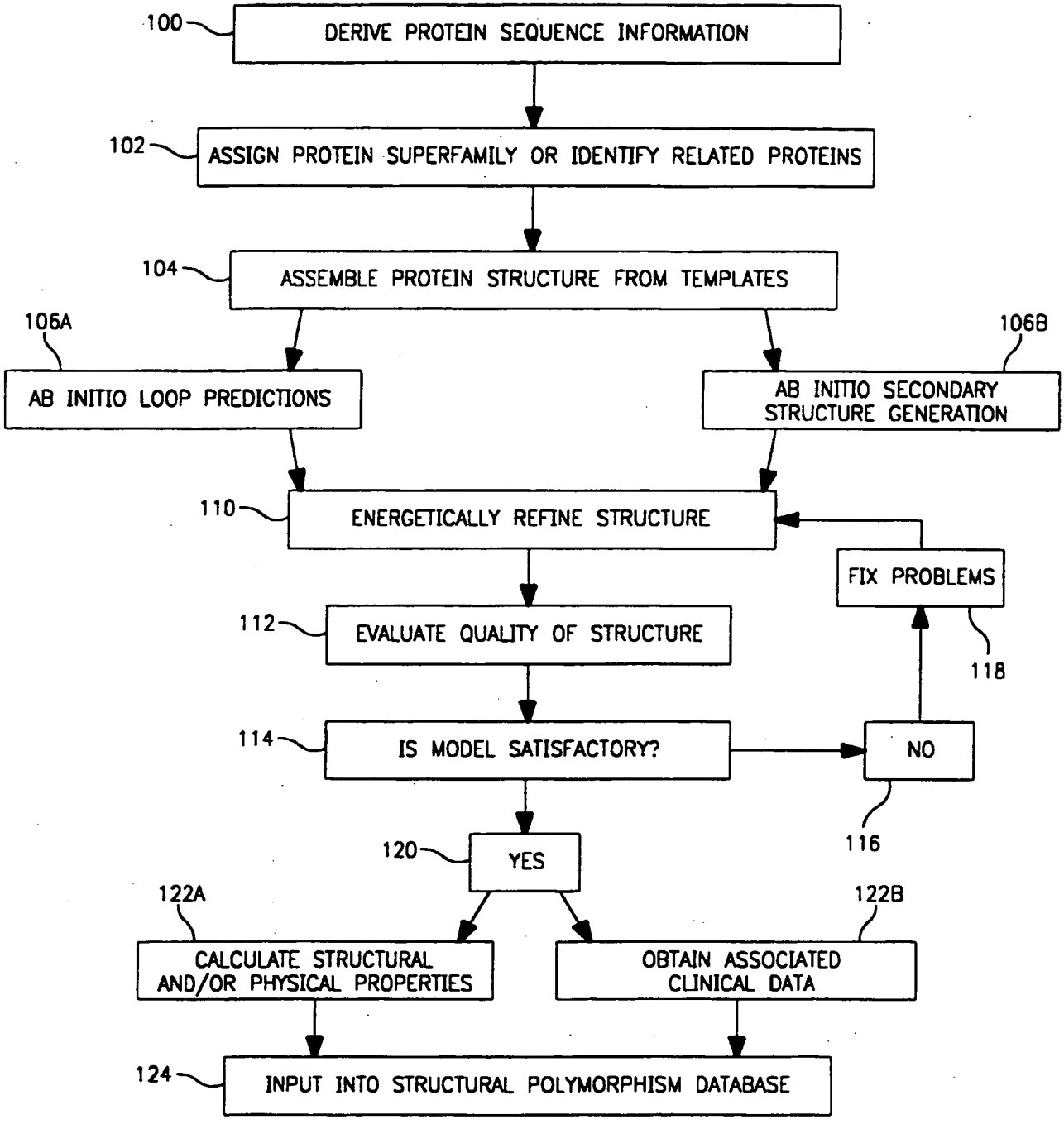

100
DERIVE PROTEIN SEQUENCE INFORMATION
102
ASSIGN PROTEIN SUPERFAMILY OR IDENTIFY RELATED PROTEINS
104
ASSEMBLE PROTEIN STRUCTURE FROM TEMPLATES
106A
AB INITIO LOOP PREDICTIONS
106B
AB INITIO SECONDARY STRUCTURE GENERATION
110
ENERGETICALLY REFINE STRUCTURE
FIX PROBLEMS
118
112
EVALUATE QUALITY OF STRUCTURE
114
IS MODEL SATISFACTORY?
NO
116
120
YES
122A
CALCULATE STRUCTURAL AND/OR PHYSICAL PROPERTIES
122B
OBTAIN ASSOCIATED CLINICAL DATA
124
INPUT INTO STRUCTURAL POLYMORPHISM DATABASE

FIG. 1

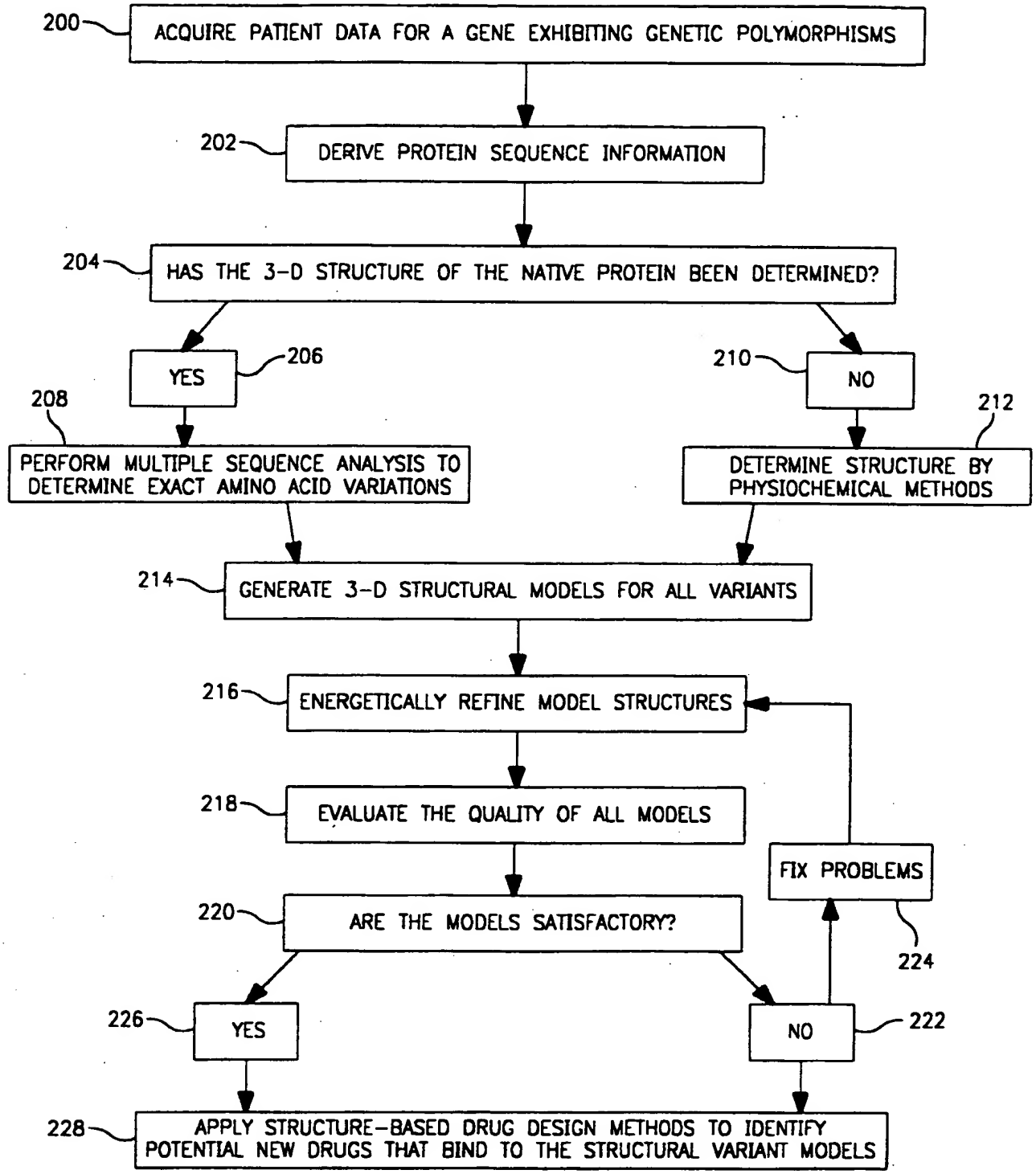

200
ACQUIRE PATIENT DATA FOR A GENE EXHIBITING GENETIC POLYMORPHISMS
202
DERIVE PROTEIN SEQUENCE INFORMATION
204
HAS THE 3-D STRUCTURE OF THE NATIVE PROTEIN BEEN DETERMINED?
YES
206
208
PERFORM MULTIPLE SEQUENCE ANALYSIS TO DETERMINE EXACT AMINO ACID VARIATIONS
210
NO
212
DETERMINE STRUCTURE BY PHYSIOCHEMICAL METHODS
214
GENERATE 3-D STRUCTURAL MODELS FOR ALL VARIANTS
216
ENERGETICALLY REFINE MODEL STRUCTURES
218
EVALUATE THE QUALITY OF ALL MODELS
220
ARE THE MODELS SATISFACTORY?
FIX PROBLEMS
224
226
YES
NO
222
228
APPLY STRUCTURE-BASED DRUG DESIGN METHODS TO IDENTIFY POTENTIAL NEW DRUGS THAT BIND TO THE STRUCTURAL VARIANT MODELS

FIG. 2

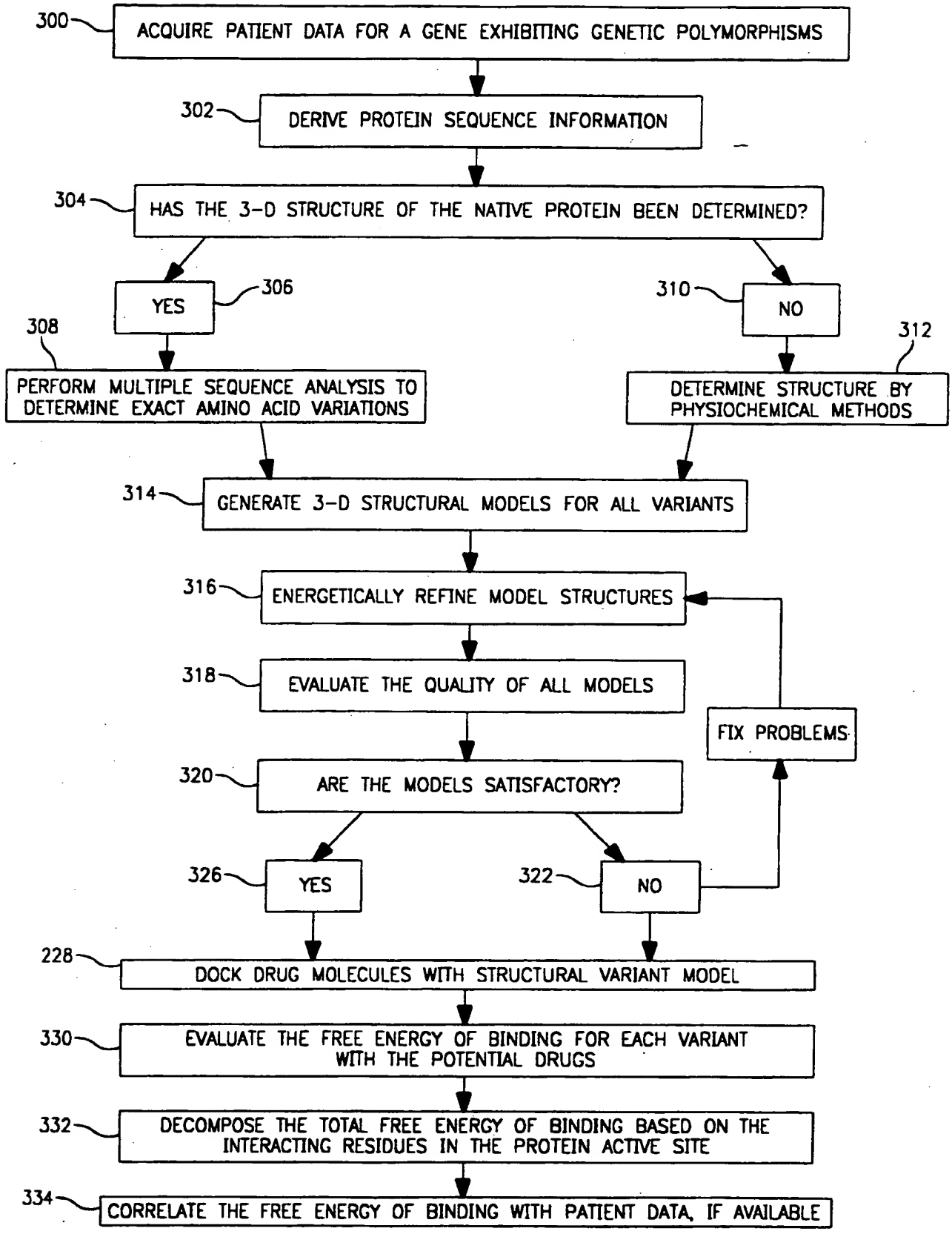
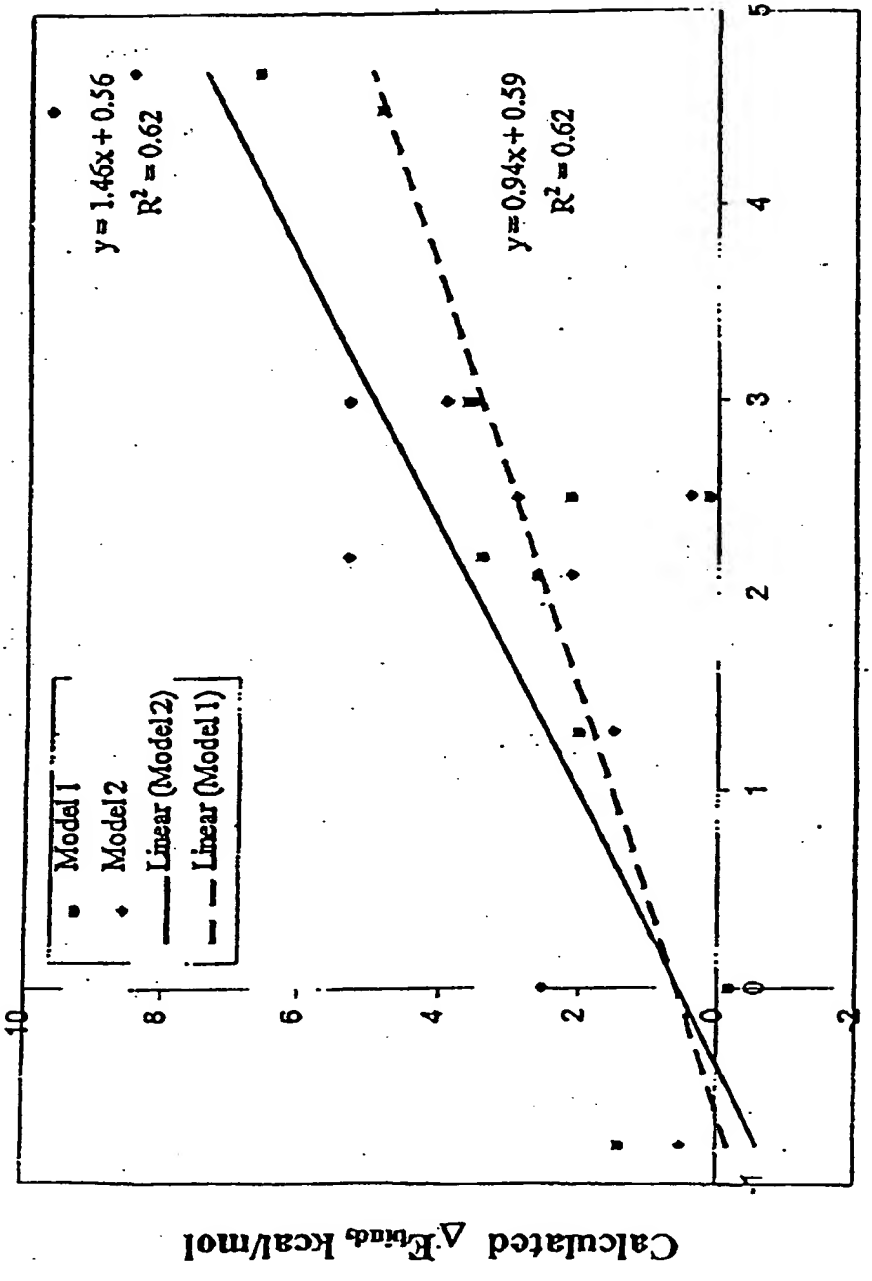

300
ACQUIRE PATIENT DATA FOR A GENE EXHIBITING GENETIC POLYMORPHISMS
302
DERIVE PROTEIN SEQUENCE INFORMATION
304
HAS THE 3-D STRUCTURE OF THE NATIVE PROTEIN BEEN DETERMINED?
YES
306
NO
310
308
PERFORM MULTIPLE SEQUENCE ANALYSIS TO DETERMINE EXACT AMINO ACID VARIATIONS
312
DETERMINE STRUCTURE BY PHYSIOCHEMICAL METHODS
314
GENERATE 3-D STRUCTURAL MODELS FOR ALL VARIANTS
316
ENERGETICALLY REFINE MODEL STRUCTURES
318
EVALUATE THE QUALITY OF ALL MODELS
FIX PROBLEMS
320
ARE THE MODELS SATISFACTORY?
326
YES
322
NO
228
DOCK DRUG MOLECULES WITH STRUCTURAL VARIANT MODEL
330
EVALUATE THE FREE ENERGY OF BINDING FOR EACH VARIANT WITH THE POTENTIAL DRUGS
332
DECOMPOSE THE TOTAL FREE ENERGY OF BINDING BASED ON THE INTERACTING RESIDUES IN THE PROTEIN ACTIVE SITE
334
CORRELATE THE FREE ENERGY OF BINDING WITH PATIENT DATA, IF AVAILABLE

FIG. 3

FIG. 4

Correlation between Experimental and Calculated Changes of Binding Energy upon Ligand Modifications in the Binding Site of NS3



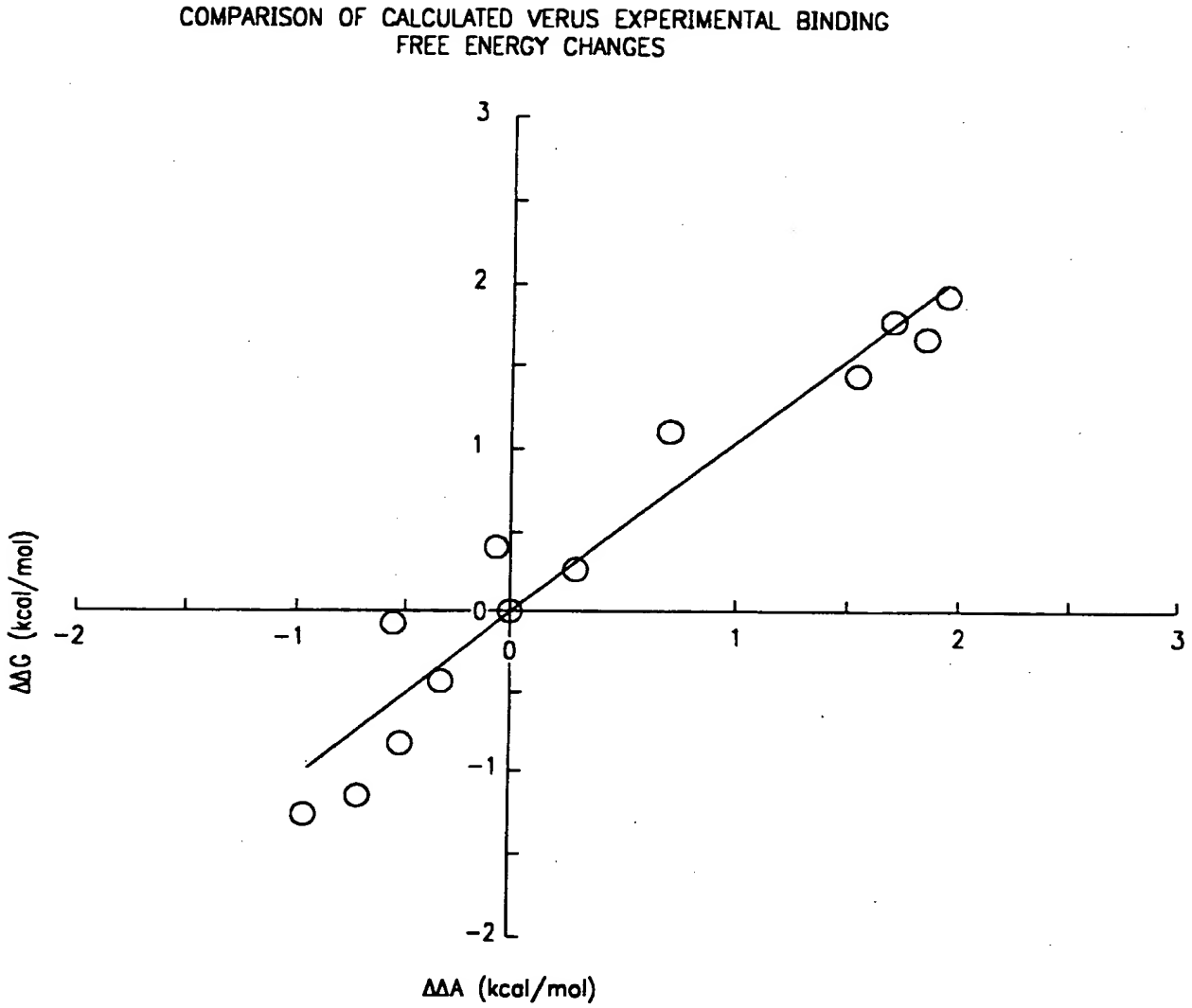
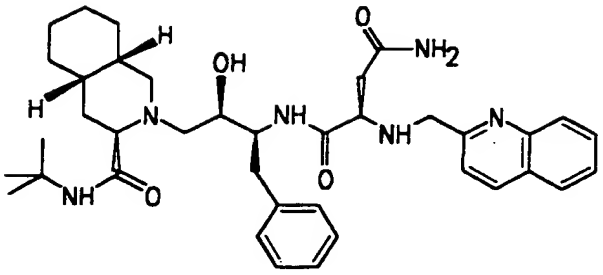
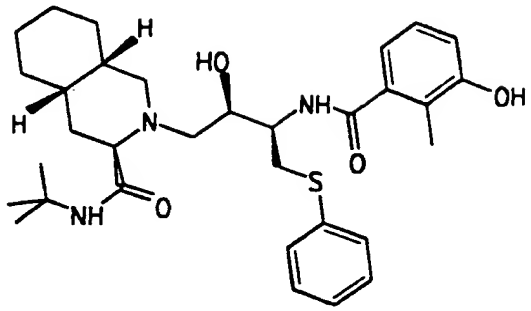

COMPARISON OF CALCULATED VERUS EXPERIMENTAL BINDING
FREE ENERGY CHANGES
ΔΔG (kcal/mol)
ΔΔA (kcal/mol)
3
2
1
0
-1
-2
-2
-1
0
1
2
3

FIG. 5

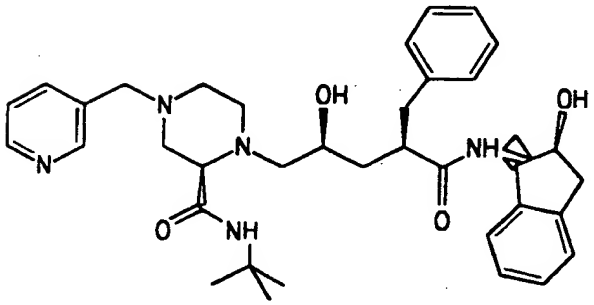
HIV PROTEASE INHIBITORS APPROVED BY FDA



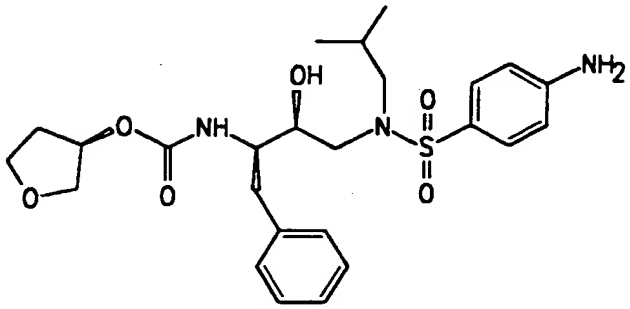
SAQUINAVIR



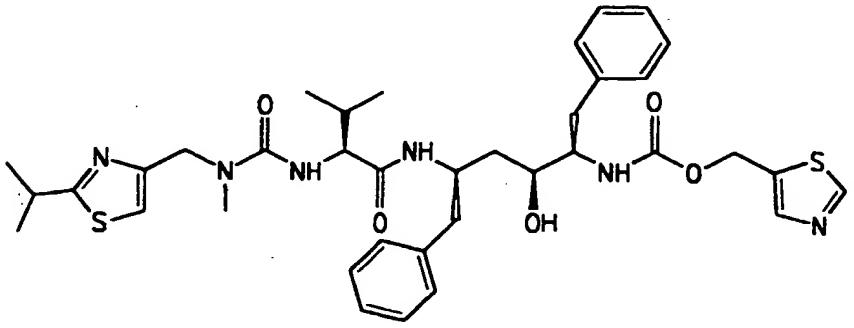
NELFINAVIR



INDINAVIR



AMPRENAVIR



RITONAVIR

FIG. 6

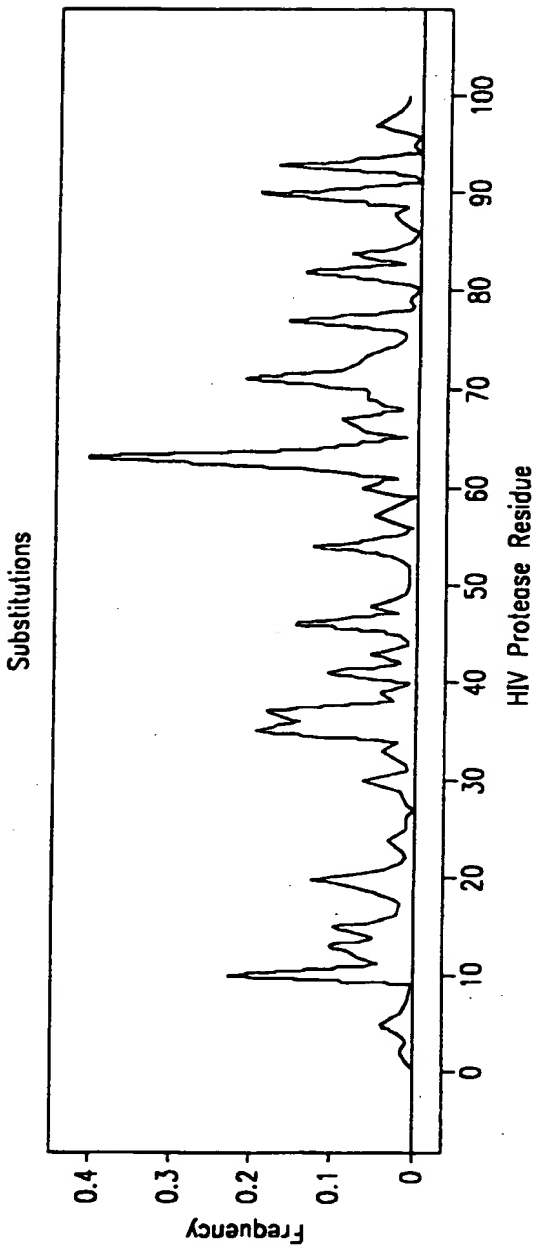

Substitutions
Frequency
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
100
HIV Protease Residue

FIG. 7

Database filename: hivpr.mdb
Number of structures: 10591
Tolerance (%) >= 1.05

| ResNum | TotOcc | TotFreq | Dist | WtAA | NumMut | MutList | NumList |
|---|---|---|---|---|---|---|---|
| 1 | 11 | 0 | 15.4 | P | 0 | | |
| 2 | 32 | 0 | 14.5 | Q | 0 | | |
| 3 | 38 | 0 | 12.1 | I | 0 | | |
| 4 | 106 | 0 | 13.0 | T | 0 | | |
| 5 | 100 | 0 | 11.3 | L | 0 | | |
| 6 | 47 | 0 | 14.3 | W | 0 | | |
| 7 | 58 | 0 | 12.8 | Q | 0 | | |
| 8 | 27 | 0 | 9.6 | R | 0 | | |
| 9 | 11 | 0 | 7.9 | P | 0 | | |
| 10 | 4004 | 37.8 | 9.2 | L | 3 | IVF | 3162 441 278 |
| 11 | 82 | 0 | 10.9 | V | 0 | | |
| 12 | 1117 | 10.5 | 13.7 | T | 5 | SEPAN | 241 185 158 155 117 |
| 13 | 1745 | 16.5 | 13.7 | I | 1 | V | 1717 |
| 14 | 646 | 6.1 | 17.0 | K | 1 | R | 623 |
| 15 | 1760 | 16.6 | 17.5 | I | 1 | V | 1709 |
| 16 | 361 | 3.4 | 20.9 | G | 1 | E | 254 |
| 17 | 56 | 0 | 22.4 | G | 0 | | |
| 18 | 242 | 2.3 | 20.5 | Q | 0 | | |
| 19 | 1340 | 12.7 | 18.3 | L | 4 | IVQT | 873 162 130 128 |
| 20 | 1549 | 14.6 | 15.4 | K | 4 | IRTM | 576 560 209 145 |
| 21 | 43 | 0 | 12.7 | E | 0 | | |
| 22 | 46 | 0 | 9.0 | A | 0 | | |
| 23 | 89 | 0 | 5.8 | L | 0 | | |
| 24 | 402 | 3.8 | 3.8 | L | 1 | I | 377 |

FIG. 8A

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 25 | 28 | 0 | 0.0 | D | 0 | | |
| 26 | 14 | 0 | 3.8 | T | 0 | | |
| 27 | 9 | 0 | 5.5 | G | 0 | | |
| 28 | 16 | 0 | 5.8 | A | 0 | | |
| 29 | 34 | 0 | 8.7 | D | 0 | | |
| 30 | 770 | 7.3 | 9.2 | D | 1 | N | 725 |
| 31 | 15 | 0 | 8.9 | T | 0 | | |
| 32 | 238 | 2.2 | 10.5 | V | 1 | I | 221 |
| 33 | 578 | 5.5 | 12.4 | L | 3 | VIF | 207 189 172 |
| 34 | 88 | 0 | 15.1 | E | 0 | | |
| 35 | 2790 | 26.3 | 18.6 | E | 1 | D | 2646 |
| 36 | 2780 | 26.2 | 20.2 | M | 2 | IV | 2549 129 |
| 37 | 3252 | 30.7 | 22.8 | N | 4 | DSET | 1253 1129 246 209 |
| 38 | 54 | 0 | 22.0 | L | 0 | | |
| 39 | 302 | 2.9 | 24.9 | P | 1 | S | 133 |
| 40 | 19 | 0 | 25.5 | G | 0 | | |
| 41 | 2249 | 21.2 | 26.0 | R | 1 | K | 2235 |
| 42 | 21 | 0 | 23.5 | W | 0 | | |
| 43 | 372 | 3.5 | 23.7 | K | 2 | TR | 166 144 |
| 44 | 12 | 0 | 22.6 | P | 0 | | |
| 45 | 208 | 2 | 20.0 | K | 1 | R | 170 |
| 46 | 2165 | 20.4 | 18.8 | M | 2 | IL | 1580 560 |
| 47 | 47 | 0 | 15.4 | I | 0 | | |
| 48 | 445 | 4.2 | 14.9 | G | 1 | V | 385 |
| 49 | 17 | 0 | 12.9 | G | 0 | | |
| 50 | 31 | 0 | 14.5 | I | 0 | | |
| 51 | 24 | 0 | 17.6 | G | 0 | | |
| 52 | 12 | 0 | 18.3 | G | 0 | | |
| 53 | 408 | 3.9 | 18.1 | F | 1 | L | 360 |

FIG. 8B

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 54 | 1661 | 15.7 | 18.0 | I | 1 | V | 1460 |
| 55 | 164 | 1.5 | 19.7 | K | 1 | R | 149 |
| 56 | 13 | 0 | 18.1 | V | 0 | | |
| 57 | 1194 | 11.3 | 19.7 | R | 1 | K | 1162 |
| 58 | 341 | 3.2 | 18.6 | Q | 1 | E | 317 |
| 59 | 20 | 0 | 19.4 | Y | 0 | | |
| 60 | 992 | 9.4 | 19.6 | D | 1 | E | 938 |
| 61 | 468 | 4.4 | 19.9 | Q | 1 | E | 285 |
| 62 | 2711 | 25.6 | 18.6 | I | 1 | V | 2685 |
| 63 | 8864 | 83.7 | 18.5 | L | 6 | PASTQH | 7245 380 321 266 226 162 |
| 64 | 2238 | 21.1 | 15.8 | I | 2 | VL | 1931 223 |
| 65 | 222 | 2.1 | 15.6 | E | 1 | D | 206 |
| 66 | 194 | 1.8 | 12.8 | I | 0 | | |
| 67 | 309 | 2.9 | 14.6 | C | 1 | S | 143 |
| 68 | 51 | 0 | 17.5 | G | 0 | | |
| 69 | 773 | 7.3 | 16.1 | H | 2 | QY | 376 206 |
| 70 | 478 | 4.5 | 17.0 | K | 1 | R | 359 |
| 71 | 3664 | 34.6 | 15.3 | A | 3 | VTI | 2301 1145 190 |
| 72 | 1494 | 14.1 | 17.2 | I | 3 | VTL | 650 409 126 |
| 73 | 1246 | 11.8 | 15.8 | G | 2 | ST | 932 185 |
| 74 | 658 | 6.2 | 15.4 | T | 2 | SA | 433 126 |
| 75 | 73 | 0 | 14.1 | V | 0 | | |
| 76 | 59 | 0 | 14.6 | L | 0 | | |
| 77 | 3533 | 33.4 | 16.1 | V | 1 | I | 3513 |
| 78 | 8 | 0 | 16.9 | G | 0 | | |
| 79 | 95 | 0 | 17.2 | P | 0 | | |
| 80 | 6 | 0 | 13.6 | T | 0 | | |
| 81 | 7 | 0 | 13.7 | P | 0 | | |
| 82 | 2208 | 20.8 | 11.0 | V | 2 | AT | 1668 284 |

FIG. 8C

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 83 | 44 | 0 | 9.7 | N | 0 | | |
| 84 | 1091 | 10.3 | 6.3 | I | 1 | V | 1073 |
| 85 | 213 | 2 | 5.7 | I | 1 | V | 198 |
| 86 | 16 | 0 | 5.3 | G | 0 | | |
| 87 | 32 | 0 | 7.3 | R | 0 | | |
| 88 | 706 | 6.7 | 10.4 | N | 2 | DS | 543 128 |
| 89 | 240 | 2.3 | 10.1 | L | 1 | M | 143 |
| 90 | 3429 | 32.4 | 8.3 | L | 1 | M | 3397 |
| 91 | 28 | 0 | 11.4 | T | 0 | | |
| 92 | 227 | 2.1 | 13.6 | Q | 1 | K | 169 |
| 93 | 3095 | 29.5 | 13.1 | I | 1 | L | 3041 |
| 94 | 15 | 0 | 13.6 | G | 0 | | |
| 95 | 100 | 0 | 10.6 | C | 0 | | |
| 96 | 6 | 0 | 11.2 | T | 0 | | |
| 97 | 83 | 0 | 10.7 | L | 0 | | |
| 98 | 44 | 0 | 14.2 | N | 0 | | |
| 99 | 35 | 0 | 16.4 | F | 0 | | |

FIG. 8D

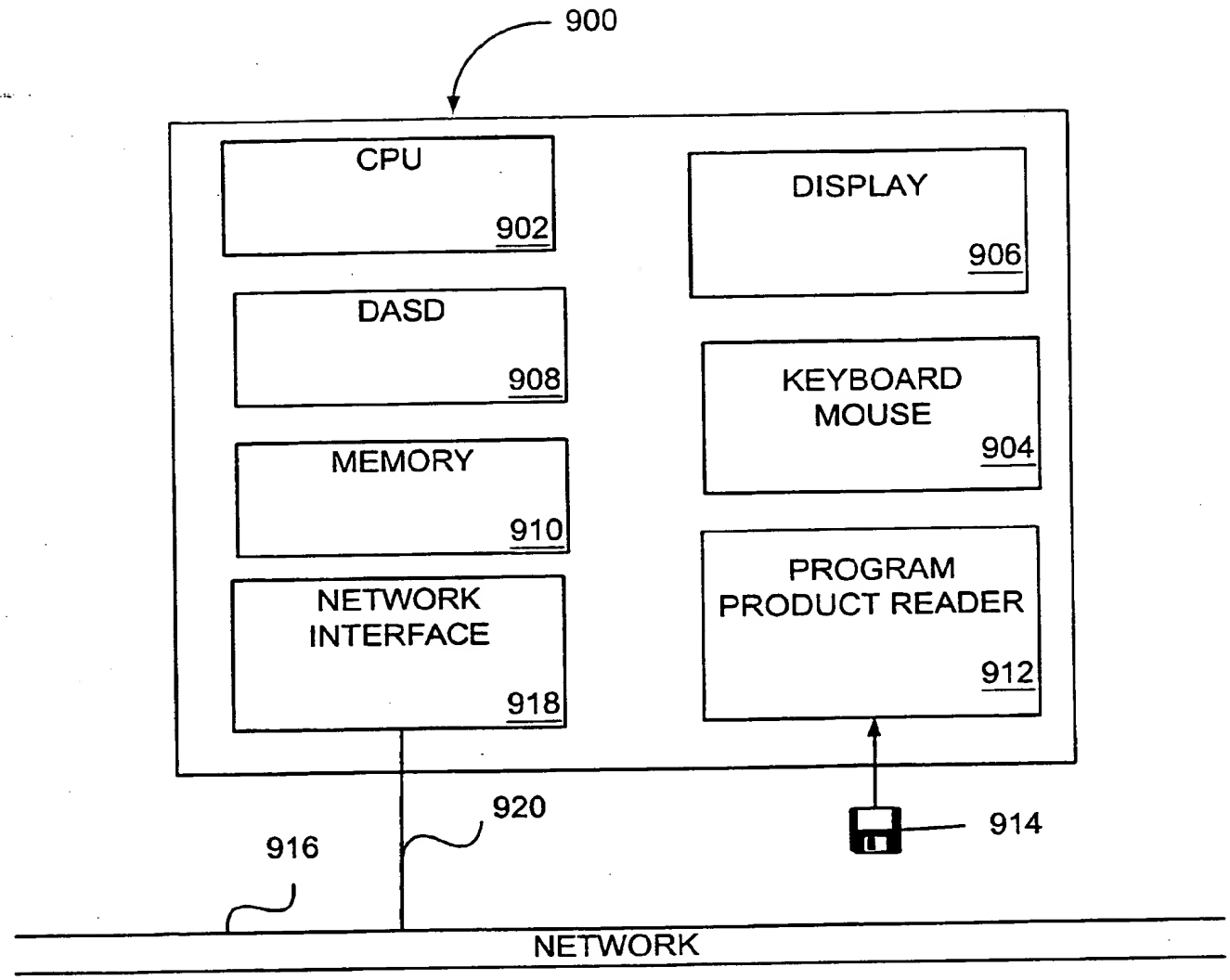

900
CPU
902
DISPLAY
906
DASD
908
KEYBOARD
MOUSE
904
MEMORY
910
PROGRAM
PRODUCT READER
912
NETWORK
INTERFACE
918
920
914
916
NETWORK

FIG. 9

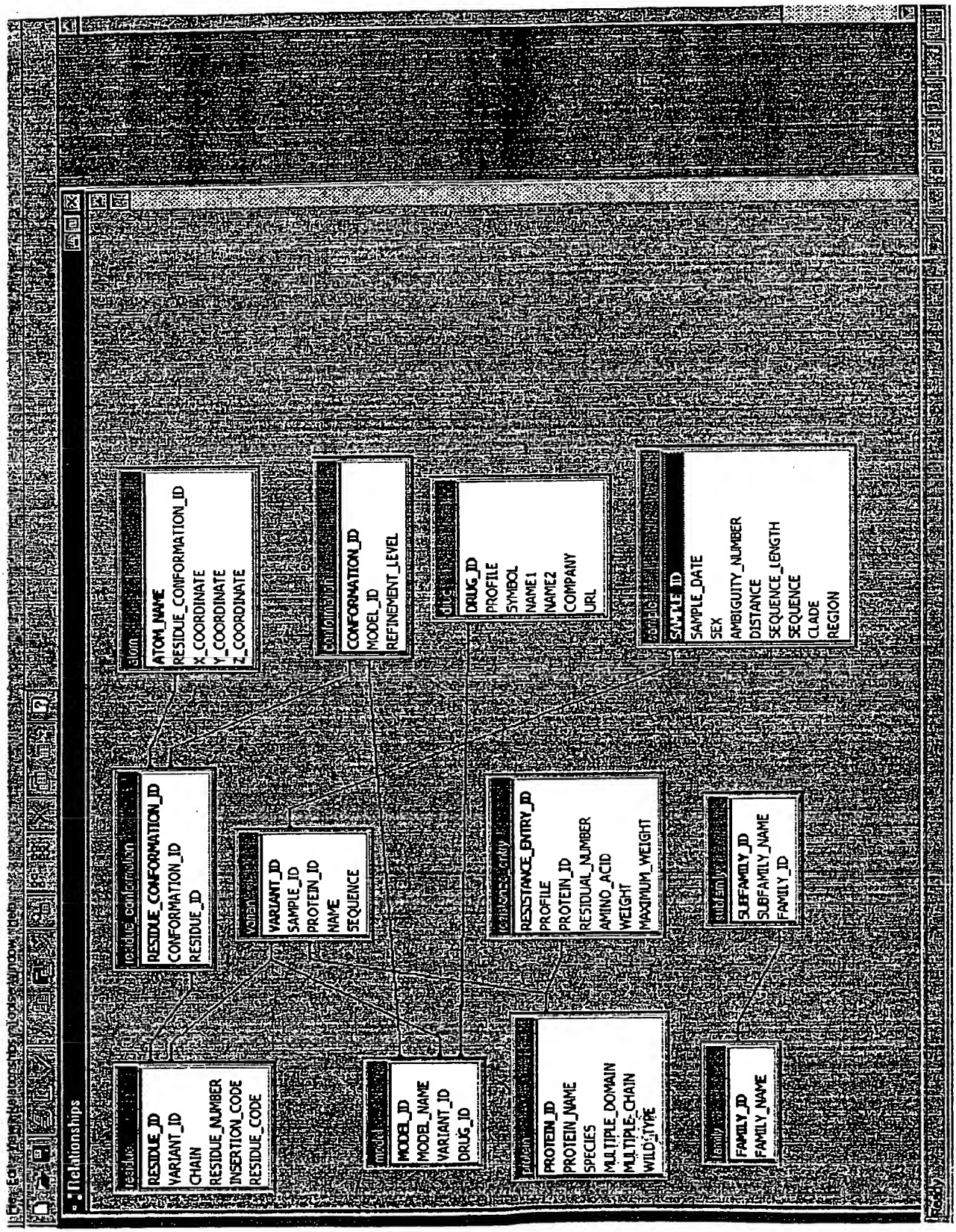

Relationships
residue
RESIDUE_ID
VARIANT_ID
CHAIN
RESIDUE_NUMBER
INSERTION_CODE
RESIDUE_CODE
residue_conformation
RESIDUE_CONFORMATION_ID
CONFORMATION_ID
RESIDUE_ID
atom
ATOM_NAME
RESIDUE_CONFORMATION_ID
X_COORDINATE
Y_COORDINATE
Z_COORDINATE
variant
VARIANT_ID
SAMPLE_ID
PROTEIN_ID
NAME
SEQUENCE
conformation
CONFORMATION_ID
MODEL_ID
REFINEMENT_LEVEL
model
MODEL_ID
MODEL_NAME
VARIANT_ID
DRUG_ID
drug
DRUG_ID
PROFILE
SYMBOL
NAME1
NAME2
COMPANY
URL
protein
PROTEIN_ID
PROTEIN_NAME
SPECIES
MULTIPLE_DOMAIN
MULTIPLE_CHAIN
WILD_TYPE
resistance_entry
RESISTANCE_ENTRY_ID
PROFILE
PROTEIN_ID
RESIDUAL_NUMBER
AMINO_ACID
WEIGHT
MAXIMUM_WEIGHT
sample
SAMPLE_ID
SAMPLE_DATE
SEX
AMBIGUITY_NUMBER
DISTANCE
SEQUENCE_LENGTH
SEQUENCE
CLADE
REGION
family
FAMILY_ID
FAMILY_NAME
subfamily
SUBFAMILY_ID
SUBFAMILY_NAME
FAMILY_ID
Ready

FIGURE 10

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1 | N | PRO | A | 1 | -3.433 | 7.956 | 34.152 |
| ATOM | 2 | CA | PRO | A | 1 | -2.653 | 6.918 | 34.784 |
| ATOM | 3 | C | PRO | A | 1 | -1.242 | 7.005 | 34.259 |
| ATOM | 4 | O | PRO | A | 1 | -0.950 | 7.638 | 33.216 |
| ATOM | 5 | CB | PRO | A | 1 | -3.281 | 5.601 | 34.262 |
| ATOM | 6 | CG | PRO | A | 1 | -4.191 | 5.995 | 33.118 |
| ATOM | 7 | CD | PRO | A | 1 | -4.547 | 7.461 | 33.339 |
| ATOM | 8 | 1H | PRO | A | 1 | -2.845 | 8.493 | 33.547 |
| ATOM | 9 | 2H | PRO | A | 1 | -3.824 | 8.552 | 34.853 |
| ATOM | 10 | N | GLN | A | 2 | -0.259 | 6.464 | 35.001 |
| ATOM | 11 | H | GLN | A | 2 | -0.475 | 6.057 | 35.889 |
| ATOM | 12 | CA | GLN | A | 2 | 1.115 | 6.443 | 34.568 |
| ATOM | 13 | C | GLN | A | 2 | 1.452 | 4.993 | 34.301 |
| ATOM | 14 | O | GLN | A | 2 | 1.379 | 4.106 | 35.173 |
| ATOM | 15 | CB | GLN | A | 2 | 2.070 | 6.966 | 35.653 |
| ATOM | 16 | CG | GLN | A | 2 | 3.549 | 6.859 | 35.240 |
| ATOM | 17 | CD | GLN | A | 2 | 4.490 | 7.744 | 36.054 |
| ATOM | 18 | OE1 | GLN | A | 2 | 4.771 | 8.888 | 35.719 |
| ATOM | 19 | NE2 | GLN | A | 2 | 4.980 | 7.190 | 37.144 |
| ATOM | 20 | 1HE2 | GLN | A | 2 | 5.605 | 7.702 | 37.734 |
| ATOM | 21 | 2HE2 | GLN | A | 2 | 4.731 | 6.253 | 37.390 |
| ATOM | 22 | N | ILE | A | 3 | 1.784 | 4.644 | 33.037 |
| ATOM | 23 | H | ILE | A | 3 | 1.876 | 5.351 | 32.336 |
| ATOM | 24 | CA | ILE | A | 3 | 2.013 | 3.257 | 32.665 |
| ATOM | 25 | C | ILE | A | 3 | 3.505 | 3.028 | 32.473 |
| ATOM | 26 | O | ILE | A | 3 | 4.242 | 3.777 | 31.787 |
| ATOM | 27 | CB | ILE | A | 3 | 1.226 | 2.944 | 31.370 |
| ATOM | 28 | CG1 | ILE | A | 3 | -0.274 | 3.239 | 31.603 |
| ATOM | 29 | CG2 | ILE | A | 3 | 1.427 | 1.480 | 30.901 |
| ATOM | 30 | CD1 | ILE | A | 3 | -1.089 | 3.219 | 30.322 |
| ATOM | 31 | N | THR | A | 4 | 4.071 | 2.032 | 33.177 |
| ATOM | 32 | H | THR | A | 4 | 3.525 | 1.525 | 33.844 |
| ATOM | 33 | CA | THR | A | 4 | 5.451 | 1.661 | 33.007 |
| ATOM | 34 | C | THR | A | 4 | 5.515 | 0.637 | 31.901 |
| ATOM | 35 | O | THR | A | 4 | 4.490 | 0.143 | 31.397 |
| ATOM | 36 | CB | THR | A | 4 | 6.051 | 1.125 | 34.324 |
| ATOM | 37 | OG1 | THR | A | 4 | 5.224 | 0.069 | 34.791 |
| ATOM | 38 | HG1 | THR | A | 4 | 5.589 | -0.299 | 35.646 |
| ATOM | 39 | CG2 | THR | A | 4 | 6.085 | 2.212 | 35.431 |
| ATOM | 40 | N | LEU | A | 5 | 6.677 | 0.281 | 31.405 |
| ATOM | 41 | H | LEU | A | 5 | 7.518 | 0.530 | 31.885 |
| ATOM | 42 | CA | LEU | A | 5 | 6.754 | -0.464 | 30.177 |
| ATOM | 43 | C | LEU | A | 5 | 7.432 | -1.813 | 30.356 |
| ATOM | 44 | O | LEU | A | 5 | 7.940 | -2.464 | 29.426 |
| ATOM | 45 | CB | LEU | A | 5 | 7.459 | 0.394 | 29.128 |
| ATOM | 46 | CG | LEU | A | 5 | 6.668 | 1.671 | 28.775 |
| ATOM | 47 | CD1 | LEU | A | 5 | 7.493 | 2.649 | 27.939 |
| ATOM | 48 | CD2 | LEU | A | 5 | 5.345 | 1.307 | 28.099 |
| ATOM | 49 | N | TRP | A | 6 | 7.420 | -2.351 | 31.594 |
| ATOM | 50 | H | TRP | A | 6 | 7.030 | -1.833 | 32.356 |
| ATOM | 51 | CA | TRP | A | 6 | 7.958 | -3.669 | 31.865 |
| ATOM | 52 | C | TRP | A | 6 | 7.071 | -4.697 | 31.204 |
| ATOM | 53 | O | TRP | A | 6 | 7.520 | -5.798 | 30.828 |
| ATOM | 54 | CB | TRP | A | 6 | 8.099 | -3.913 | 33.367 |
| ATOM | 55 | CG | TRP | A | 6 | 9.041 | -2.974 | 34.070 |

FIG. 11A

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 56 | CD1 | TRP | A | 6 | 8.745 | -1.769 | 34.646 |
| ATOM | 57 | CD2 | TRP | A | 6 | 10.449 | -3.171 | 34.273 |
| ATOM | 58 | NE1 | TRP | A | 6 | 9.875 | -1.209 | 35.190 |
| ATOM | 59 | HE1 | TRP | A | 6 | 9.930 | -0.332 | 35.668 |
| ATOM | 60 | CE2 | TRP | A | 6 | 10.932 | -2.048 | 34.974 |
| ATOM | 61 | CE3 | TRP | A | 6 | 11.334 | -4.190 | 33.924 |
| ATOM | 62 | CZ2 | TRP | A | 6 | 12.257 | -1.917 | 35.333 |
| ATOM | 63 | CZ3 | TRP | A | 6 | 12.650 | -4.065 | 34.278 |
| ATOM | 64 | CH2 | TRP | A | 6 | 13.106 | -2.942 | 34.974 |
| ATOM | 65 | N | GLN | A | 7 | 5.773 | -4.448 | 30.973 |
| ATOM | 66 | H | GLN | A | 7 | 5.354 | -3.619 | 31.343 |
| ATOM | 67 | CA | GLN | A | 7 | 4.952 | -5.339 | 30.205 |
| ATOM | 68 | C | GLN | A | 7 | 4.438 | -4.569 | 29.033 |
| ATOM | 69 | O | GLN | A | 7 | 4.433 | -3.321 | 29.000 |
| ATOM | 70 | CB | GLN | A | 7 | 3.712 | -5.693 | 30.969 |
| ATOM | 71 | CG | GLN | A | 7 | 4.015 | -6.467 | 32.210 |
| ATOM | 72 | CD | GLN | A | 7 | 2.734 | -6.678 | 32.917 |
| ATOM | 73 | OE1 | GLN | A | 7 | 2.053 | -7.681 | 32.712 |
| ATOM | 74 | NE2 | GLN | A | 7 | 2.356 | -5.682 | 33.736 |
| ATOM | 75 | 1HE2 | GLN | A | 7 | 1.501 | -5.748 | 34.251 |
| ATOM | 76 | 2HE2 | GLN | A | 7 | 2.926 | -4.867 | 33.837 |
| ATOM | 77 | N | ARG | A | 8 | 3.777 | -5.239 | 28.078 |
| ATOM | 78 | H | ARG | A | 8 | 3.688 | -6.233 | 28.142 |
| ATOM | 79 | CA | ARG | A | 8 | 3.183 | -4.568 | 26.948 |
| ATOM | 80 | C | ARG | A | 8 | 2.117 | -3.648 | 27.461 |
| ATOM | 81 | O | ARG | A | 8 | 1.333 | -3.965 | 28.387 |
| ATOM | 82 | CB | ARG | A | 8 | 2.574 | -5.555 | 25.975 |
| ATOM | 83 | CG | ARG | A | 8 | 3.532 | -6.593 | 25.437 |
| ATOM | 84 | CD | ARG | A | 8 | 2.842 | -7.610 | 24.579 |
| ATOM | 85 | NE | ARG | A | 8 | 3.787 | -8.487 | 23.900 |
| ATOM | 86 | HE | ARG | A | 8 | 4.762 | -8.279 | 23.982 |
| ATOM | 87 | CZ | ARG | A | 8 | 3.405 | -9.541 | 23.185 |
| ATOM | 88 | NH1 | ARG | A | 8 | 2.125 | -9.871 | 23.052 |
| ATOM | 89 | 2HH1 | ARG | A | 8 | 1.418 | -9.321 | 23.496 |
| ATOM | 90 | 1HH1 | ARG | A | 8 | 1.869 | -10.670 | 22.508 |
| ATOM | 91 | NH2 | ARG | A | 8 | 4.332 | -10.286 | 22.589 |
| ATOM | 92 | 1HH2 | ARG | A | 8 | 4.062 | -11.082 | 22.048 |
| ATOM | 93 | 2HH2 | ARG | A | 8 | 5.299 | -10.050 | 22.682 |
| ATOM | 94 | N | PRO | A | 9 | 1.990 | -2.428 | 26.938 |
| ATOM | 95 | CA | PRO | A | 9 | 1.001 | -1.462 | 27.440 |
| ATOM | 96 | C | PRO | A | 9 | -0.365 | -1.697 | 26.821 |
| ATOM | 97 | O | PRO | A | 9 | -0.918 | -0.935 | 26.008 |
| ATOM | 98 | CB | PRO | A | 9 | 1.572 | -0.112 | 27.041 |
| ATOM | 99 | CG | PRO | A | 9 | 2.553 | -0.404 | 25.931 |
| ATOM | 100 | CD | PRO | A | 9 | 3.024 | -1.820 | 26.084 |
| ATOM | 101 | N | LEU | A | 10 | -1.028 | -2.803 | 27.227 |
| ATOM | 102 | H | LEU | A | 10 | -0.616 | -3.404 | 27.912 |
| ATOM | 103 | CA | LEU | A | 10 | -2.319 | -3.143 | 26.698 |
| ATOM | 104 | C | LEU | A | 10 | -3.390 | -2.565 | 27.591 |
| ATOM | 105 | O | LEU | A | 10 | -3.336 | -2.632 | 28.831 |
| ATOM | 106 | CB | LEU | A | 10 | -2.451 | -4.651 | 26.709 |
| ATOM | 107 | CG | LEU | A | 10 | -1.483 | -5.316 | 25.756 |
| ATOM | 108 | CD1 | LEU | A | 10 | -1.159 | -6.740 | 26.212 |
| ATOM | 109 | CD2 | LEU | A | 10 | -2.083 | -5.262 | 24.322 |
| ATOM | 110 | N | VAL | A | 11 | -4.447 | -1.952 | 27.033 |
| ATOM | 111 | H | VAL | A | 11 | -4.507 | -1.875 | 26.038 |

FIG. 11B

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 112 | CA | VAL | A | 11 | -5.506 | -1.398 | 27.835 |
| ATOM | 113 | C | VAL | A | 11 | -6.827 | -1.857 | 27.268 |
| ATOM | 114 | O | VAL | A | 11 | -6.924 | -2.490 | 26.198 |
| ATOM | 115 | CB | VAL | A | 11 | -5.420 | 0.143 | 27.897 |
| ATOM | 116 | CG1 | VAL | A | 11 | -4.117 | 0.595 | 28.551 |
| ATOM | 117 | CG2 | VAL | A | 11 | -5.549 | 0.787 | 26.497 |
| ATOM | 118 | N | THR | A | 12 | -7.954 | -1.592 | 27.978 |
| ATOM | 119 | H | THR | A | 12 | -7.884 | -1.141 | 28.868 |
| ATOM | 120 | CA | THR | A | 12 | -9.301 | -1.942 | 27.496 |
| ATOM | 121 | C | THR | A | 12 | -9.889 | -0.726 | 26.795 |
| ATOM | 122 | O | THR | A | 12 | -9.856 | 0.436 | 27.247 |
| ATOM | 123 | CB | THR | A | 12 | -10.225 | -2.385 | 28.659 |
| ATOM | 124 | OG1 | THR | A | 12 | -9.596 | -3.458 | 29.338 |
| ATOM | 125 | HG1 | THR | A | 12 | -10.170 | -3.766 | 30.096 |
| ATOM | 126 | CG2 | THR | A | 12 | -11.579 | -2.895 | 28.156 |
| ATOM | 127 | N | ILE | A | 13 | -10.449 | -0.932 | 25.594 |
| ATOM | 128 | H | ILE | A | 13 | -10.409 | -1.841 | 25.178 |
| ATOM | 129 | CA | ILE | A | 13 | -11.112 | 0.133 | 24.882 |
| ATOM | 130 | C | ILE | A | 13 | -12.553 | -0.292 | 24.693 |
| ATOM | 131 | O | ILE | A | 13 | -12.935 | -1.469 | 24.821 |
| ATOM | 132 | CB | ILE | A | 13 | -10.432 | 0.364 | 23.511 |
| ATOM | 133 | CG1 | ILE | A | 13 | -10.466 | -0.896 | 22.628 |
| ATOM | 134 | CG2 | ILE | A | 13 | -8.986 | 0.806 | 23.747 |
| ATOM | 135 | CD1 | ILE | A | 13 | -9.755 | -0.745 | 21.294 |
| ATOM | 136 | N | LYS | A | 14 | -13.470 | 0.658 | 24.438 |
| ATOM | 137 | H | LYS | A | 14 | -13.209 | 1.622 | 24.481 |
| ATOM | 138 | CA | LYS | A | 14 | -14.838 | 0.330 | 24.100 |
| ATOM | 139 | C | LYS | A | 14 | -15.088 | 0.877 | 22.719 |
| ATOM | 140 | O | LYS | A | 14 | -14.859 | 2.059 | 22.375 |
| ATOM | 141 | CB | LYS | A | 14 | -15.855 | 0.916 | 25.099 |
| ATOM | 142 | CG | LYS | A | 14 | -17.325 | 0.518 | 24.864 |
| ATOM | 143 | CD | LYS | A | 14 | -18.078 | 0.146 | 26.166 |
| ATOM | 144 | CE | LYS | A | 14 | -18.826 | 1.342 | 26.810 |
| ATOM | 145 | NZ | LYS | A | 14 | -19.316 | 0.929 | 28.173 |
| ATOM | 146 | 1HZ | LYS | A | 14 | -19.801 | 1.693 | 28.599 |
| ATOM | 147 | 3HZ | LYS | A | 14 | -18.536 | 0.670 | 28.743 |
| ATOM | 148 | 2HZ | LYS | A | 14 | -19.936 | 0.150 | 28.082 |
| ATOM | 149 | N | ILE | A | 15 | -15.535 | 0.005 | 21.798 |
| ATOM | 150 | H | ILE | A | 15 | -15.806 | -0.916 | 22.078 |
| ATOM | 151 | CA | ILE | A | 15 | -15.642 | 0.347 | 20.400 |
| ATOM | 152 | C | ILE | A | 15 | -16.894 | -0.328 | 19.887 |
| ATOM | 153 | O | ILE | A | 15 | -17.115 | -1.542 | 20.041 |
| ATOM | 154 | CB | ILE | A | 15 | -14.382 | -0.132 | 19.639 |
| ATOM | 155 | CG1 | ILE | A | 15 | -14.478 | 0.148 | 18.125 |
| ATOM | 156 | CG2 | ILE | A | 15 | -14.082 | -1.623 | 19.880 |
| ATOM | 157 | CD1 | ILE | A | 15 | -14.237 | 1.603 | 17.796 |
| ATOM | 158 | N | GLY | A | 16 | -17.843 | 0.435 | 19.308 |
| ATOM | 159 | H | GLY | A | 16 | -17.720 | 1.426 | 19.260 |
| ATOM | 160 | CA | GLY | A | 16 | -19.053 | -0.143 | 18.745 |
| ATOM | 161 | C | GLY | A | 16 | -19.897 | -0.817 | 19.789 |
| ATOM | 162 | O | GLY | A | 16 | -20.774 | -1.668 | 19.516 |
| ATOM | 163 | N | GLY | A | 17 | -19.712 | -0.493 | 21.088 |
| ATOM | 164 | H | GLY | A | 17 | -19.038 | 0.204 | 21.334 |
| ATOM | 165 | CA | GLY | A | 17 | -20.464 | -1.126 | 22.160 |
| ATOM | 166 | C | GLY | A | 17 | -19.718 | -2.335 | 22.653 |
| ATOM | 167 | O | GLY | A | 17 | -20.147 | -3.098 | 23.540 |

FIG. 11C

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 168 | N | GLN | A | 18 | -18.507 | -2.591 | 22.121 |
| ATOM | 169 | H | GLN | A | 18 | -18.059 | -1.900 | 21.554 |
| ATOM | 170 | CA | GLN | A | 18 | -17.806 | -3.830 | 22.326 |
| ATOM | 171 | C | GLN | A | 18 | -16.552 | -3.549 | 23.123 |
| ATOM | 172 | O | GLN | A | 18 | -15.887 | -2.508 | 22.945 |
| ATOM | 173 | CB | GLN | A | 18 | -17.393 | -4.294 | 20.928 |
| ATOM | 174 | CG | GLN | A | 18 | -16.911 | -5.734 | 20.788 |
| ATOM | 175 | CD | GLN | A | 18 | -18.018 | -6.728 | 20.613 |
| ATOM | 176 | OE1 | GLN | A | 18 | -19.131 | -6.574 | 21.152 |
| ATOM | 177 | NE2 | GLN | A | 18 | -17.722 | -7.773 | 19.857 |
| ATOM | 178 | 1HE2 | GLN | A | 18 | -18.404 | -8.484 | 19.689 |
| ATOM | 179 | 2HE2 | GLN | A | 18 | -16.814 | -7.860 | 19.448 |
| ATOM | 180 | N | LEU | A | 19 | -16.133 | -4.397 | 24.087 |
| ATOM | 181 | H | LEU | A | 19 | -16.682 | -5.202 | 24.312 |
| ATOM | 182 | CA | LEU | A | 19 | -14.909 | -4.178 | 24.808 |
| ATOM | 183 | C | LEU | A | 19 | -13.799 | -4.912 | 24.090 |
| ATOM | 184 | O | LEU | A | 19 | -13.989 | -6.018 | 23.558 |
| ATOM | 185 | CB | LEU | A | 19 | -14.982 | -4.714 | 26.254 |
| ATOM | 186 | CG | LEU | A | 19 | -15.490 | -3.778 | 27.374 |
| ATOM | 187 | CD1 | LEU | A | 19 | -16.392 | -2.639 | 26.856 |
| ATOM | 188 | CD2 | LEU | A | 19 | -16.208 | -4.516 | 28.465 |
| ATOM | 189 | N | LYS | A | 20 | -12.603 | -4.372 | 23.978 |
| ATOM | 190 | H | LYS | A | 20 | -12.442 | -3.448 | 24.324 |
| ATOM | 191 | CA | LYS | A | 20 | -11.507 | -5.082 | 23.365 |
| ATOM | 192 | C | LYS | A | 20 | -10.266 | -4.618 | 24.062 |
| ATOM | 193 | O | LYS | A | 20 | -10.228 | -3.611 | 24.816 |
| ATOM | 194 | CB | LYS | A | 20 | -11.397 | -4.798 | 21.875 |
| ATOM | 195 | CG | LYS | A | 20 | -12.558 | -5.356 | 21.100 |
| ATOM | 196 | CD | LYS | A | 20 | -12.537 | -4.988 | 19.615 |
| ATOM | 197 | CE | LYS | A | 20 | -13.414 | -5.958 | 18.827 |
| ATOM | 198 | NZ | LYS | A | 20 | -12.681 | -7.208 | 18.639 |
| ATOM | 199 | 1HZ | LYS | A | 20 | -13.247 | -7.852 | 18.123 |
| ATOM | 200 | 3HZ | LYS | A | 20 | -12.458 | -7.601 | 19.531 |
| ATOM | 201 | 2HZ | LYS | A | 20 | -11.837 | -7.027 | 18.134 |
| ATOM | 202 | N | GLU | A | 21 | -9.150 | -5.357 | 23.893 |
| ATOM | 203 | H | GLU | A | 21 | -9.185 | -6.188 | 23.338 |
| ATOM | 204 | CA | GLU | A | 21 | -7.890 | -4.997 | 24.486 |
| ATOM | 205 | C | GLU | A | 21 | -7.001 | -4.462 | 23.390 |
| ATOM | 206 | O | GLU | A | 21 | -6.970 | -4.992 | 22.258 |
| ATOM | 207 | CB | GLU | A | 21 | -7.268 | -6.260 | 25.051 |
| ATOM | 208 | CG | GLU | A | 21 | -5.835 | -6.140 | 25.480 |
| ATOM | 209 | CD | GLU | A | 21 | -5.405 | -7.352 | 26.275 |
| ATOM | 210 | OE1 | GLU | A | 21 | -5.624 | -7.343 | 27.508 |
| ATOM | 211 | OE2 | GLU | A | 21 | -4.852 | -8.309 | 25.684 |
| ATOM | 212 | N | ALA | A | 22 | -6.239 | -3.369 | 23.595 |
| ATOM | 213 | H | ALA | A | 22 | -6.223 | -2.938 | 24.497 |
| ATOM | 214 | CA | ALA | A | 22 | -5.419 | -2.781 | 22.520 |
| ATOM | 215 | C | ALA | A | 22 | -4.138 | -2.255 | 23.114 |
| ATOM | 216 | O | ALA | A | 22 | -3.985 | -1.914 | 24.314 |
| ATOM | 217 | CB | ALA | A | 22 | -6.134 | -1.657 | 21.821 |
| ATOM | 218 | N | LEU | A | 23 | -3.121 | -2.091 | 22.240 |
| ATOM | 219 | H | LEU | A | 23 | -3.279 | -2.236 | 21.263 |
| ATOM | 220 | CA | LEU | A | 23 | -1.797 | -1.712 | 22.640 |
| ATOM | 221 | C | LEU | A | 23 | -1.660 | -0.230 | 22.443 |
| ATOM | 222 | O | LEU | A | 23 | -2.020 | 0.349 | 21.402 |
| ATOM | 223 | CB | LEU | A | 23 | -0.814 | -2.486 | 21.732 |

FIG. 1 ID

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 224 | CG | LEU | A | 23 | 0.705 | -2.448 | 21.991 |
| ATOM | 225 | CD1 | LEU | A | 23 | 1.088 | -3.400 | 23.124 |
| ATOM | 226 | CD2 | LEU | A | 23 | 1.462 | -2.878 | 20.708 |
| ATOM | 227 | N | LEU | A | 24 | -1.192 | 0.530 | 23.463 |
| ATOM | 228 | H | LEU | A | 24 | -1.015 | 0.110 | 24.353 |
| ATOM | 229 | CA | LEU | A | 24 | -0.935 | 1.952 | 23.305 |
| ATOM | 230 | C | LEU | A | 24 | 0.403 | 2.089 | 22.609 |
| ATOM | 231 | O | LEU | A | 24 | 1.471 | 1.717 | 23.130 |
| ATOM | 232 | CB | LEU | A | 24 | -0.921 | 2.609 | 24.681 |
| ATOM | 233 | CG | LEU | A | 24 | -2.220 | 2.492 | 25.477 |
| ATOM | 234 | CD1 | LEU | A | 24 | -2.063 | 3.291 | 26.772 |
| ATOM | 235 | CD2 | LEU | A | 24 | -3.419 | 3.000 | 24.691 |
| ATOM | 236 | N | ASP | A | 25 | 0.454 | 2.590 | 21.397 |
| ATOM | 237 | H | ASP | A | 25 | -0.334 | 3.085 | 21.032 |
| ATOM | 238 | CA | ASP | A | 25 | 1.642 | 2.423 | 20.605 |
| ATOM | 239 | C | ASP | A | 25 | 2.130 | 3.750 | 20.059 |
| ATOM | 240 | O | ASP | A | 25 | 1.568 | 4.320 | 19.110 |
| ATOM | 241 | CB | ASP | A | 25 | 1.263 | 1.435 | 19.486 |
| ATOM | 242 | CG | ASP | A | 25 | 2.428 | 1.051 | 18.561 |
| ATOM | 243 | OD1 | ASP | A | 25 | 3.546 | 1.540 | 18.729 |
| ATOM | 244 | OD2 | ASP | A | 25 | 2.164 | 0.241 | 17.658 |
| ATOM | 245 | N | THR | A | 26 | 3.203 | 4.337 | 20.605 |
| ATOM | 246 | H | THR | A | 26 | 3.694 | 3.880 | 21.346 |
| ATOM | 247 | CA | THR | A | 26 | 3.691 | 5.652 | 20.144 |
| ATOM | 248 | C | THR | A | 26 | 4.397 | 5.583 | 18.778 |
| ATOM | 249 | O | THR | A | 26 | 4.642 | 6.587 | 18.079 |
| ATOM | 250 | CB | THR | A | 26 | 4.596 | 6.219 | 21.217 |
| ATOM | 251 | OG1 | THR | A | 26 | 5.716 | 5.324 | 21.386 |
| ATOM | 252 | HG1 | THR | A | 26 | 6.332 | 5.676 | 22.091 |
| ATOM | 253 | CG2 | THR | A | 26 | 3.878 | 6.320 | 22.577 |
| ATOM | 254 | N | GLY | A | 27 | 4.757 | 4.377 | 18.298 |
| ATOM | 255 | H | GLY | A | 27 | 4.526 | 3.550 | 18.811 |
| ATOM | 256 | CA | GLY | A | 27 | 5.481 | 4.233 | 17.040 |
| ATOM | 257 | C | GLY | A | 27 | 4.520 | 4.190 | 15.886 |
| ATOM | 258 | O | GLY | A | 27 | 4.908 | 4.242 | 14.696 |
| ATOM | 259 | N | ALA | A | 28 | 3.197 | 4.084 | 16.117 |
| ATOM | 260 | H | ALA | A | 28 | 2.856 | 4.091 | 17.057 |
| ATOM | 261 | CA | ALA | A | 28 | 2.213 | 3.955 | 15.018 |
| ATOM | 262 | C | ALA | A | 28 | 1.598 | 5.299 | 14.750 |
| ATOM | 263 | O | ALA | A | 28 | 1.062 | 5.982 | 15.650 |
| ATOM | 264 | CB | ALA | A | 28 | 1.117 | 2.980 | 15.390 |
| ATOM | 265 | N | ASP | A | 29 | 1.503 | 5.744 | 13.490 |
| ATOM | 266 | H | ASP | A | 29 | 1.912 | 5.216 | 12.746 |
| ATOM | 267 | CA | ASP | A | 29 | 0.810 | 6.984 | 13.213 |
| ATOM | 268 | C | ASP | A | 29 | -0.666 | 6.724 | 13.327 |
| ATOM | 269 | O | ASP | A | 29 | -1.488 | 7.637 | 13.568 |
| ATOM | 270 | CB | ASP | A | 29 | 1.009 | 7.433 | 11.775 |
| ATOM | 271 | CG | ASP | A | 29 | 2.439 | 7.882 | 11.412 |
| ATOM | 272 | OD1 | ASP | A | 29 | 3.360 | 7.856 | 12.269 |
| ATOM | 273 | OD2 | ASP | A | 29 | 2.606 | 8.253 | 10.252 |
| ATOM | 274 | N | ASP | A | 30 | -1.143 | 5.517 | 12.990 |
| ATOM | 275 | H | ASP | A | 30 | -0.508 | 4.769 | 12.800 |
| ATOM | 276 | CA | ASP | A | 30 | -2.579 | 5.245 | 12.887 |
| ATOM | 277 | C | ASP | A | 30 | -3.057 | 4.208 | 13.867 |
| ATOM | 278 | O | ASP | A | 30 | -2.284 | 3.483 | 14.546 |
| ATOM | 279 | CB | ASP | A | 30 | -2.896 | 4.758 | 11.456 |

FIG. 1E

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 280 | CG | ASP | A | 30 | -2.495 | 5.768 | 10.425 |
| ATOM | 281 | OD1 | ASP | A | 30 | -3.067 | 6.871 | 10.423 |
| ATOM | 282 | OD2 | ASP | A | 30 | -1.596 | 5.494 | 9.618 |
| ATOM | 283 | N | THR | A | 31 | -4.393 | 4.076 | 14.002 |
| ATOM | 284 | H | THR | A | 31 | -5.004 | 4.700 | 13.515 |
| ATOM | 285 | CA | THR | A | 31 | -5.059 | 3.062 | 14.829 |
| ATOM | 286 | C | THR | A | 31 | -5.565 | 1.967 | 13.913 |
| ATOM | 287 | O | THR | A | 31 | -6.223 | 2.169 | 12.870 |
| ATOM | 288 | CB | THR | A | 31 | -6.212 | 3.725 | 15.566 |
| ATOM | 289 | OG1 | THR | A | 31 | -5.668 | 4.667 | 16.474 |
| ATOM | 290 | HG1 | THR | A | 31 | -6.403 | 5.122 | 16.976 |
| ATOM | 291 | CG2 | THR | A | 31 | -7.044 | 2.702 | 16.389 |
| ATOM | 292 | N | VAL | A | 32 | -5.187 | 0.713 | 14.235 |
| ATOM | 293 | H | VAL | A | 32 | -4.649 | 0.555 | 15.063 |
| ATOM | 294 | CA | VAL | A | 32 | -5.517 | -0.462 | 13.437 |
| ATOM | 295 | C | VAL | A | 32 | -6.092 | -1.506 | 14.365 |
| ATOM | 296 | O | VAL | A | 32 | -5.502 | -1.957 | 15.365 |
| ATOM | 297 | CB | VAL | A | 32 | -4.260 | -1.064 | 12.757 |
| ATOM | 298 | CG1 | VAL | A | 32 | -4.667 | -2.136 | 11.735 |
| ATOM | 299 | CG2 | VAL | A | 32 | -3.422 | 0.017 | 12.032 |
| ATOM | 300 | N | LEU | A | 33 | -7.352 | -1.923 | 14.119 |
| ATOM | 301 | H | LEU | A | 33 | -7.867 | -1.523 | 13.361 |
| ATOM | 302 | CA | LEU | A | 33 | -7.982 | -2.940 | 14.929 |
| ATOM | 303 | C | LEU | A | 33 | -8.174 | -4.203 | 14.107 |
| ATOM | 304 | O | LEU | A | 33 | -8.268 | -4.247 | 12.853 |
| ATOM | 305 | CB | LEU | A | 33 | -9.336 | -2.477 | 15.408 |
| ATOM | 306 | CG | LEU | A | 33 | -9.292 | -1.149 | 16.127 |
| ATOM | 307 | CD1 | LEU | A | 33 | -10.710 | -0.747 | 16.485 |
| ATOM | 308 | CD2 | LEU | A | 33 | -8.348 | -1.139 | 17.347 |
| ATOM | 309 | N | GLU | A | 34 | -8.296 | -5.319 | 14.782 |
| ATOM | 310 | H | GLU | A | 34 | -8.244 | -5.302 | 15.780 |
| ATOM | 311 | CA | GLU | A | 34 | -8.503 | -6.551 | 14.086 |
| ATOM | 312 | C | GLU | A | 34 | -9.909 | -6.549 | 13.510 |
| ATOM | 313 | O | GLU | A | 34 | -10.808 | -5.717 | 13.795 |
| ATOM | 314 | CB | GLU | A | 34 | -8.265 | -7.750 | 15.010 |
| ATOM | 315 | CG | GLU | A | 34 | -9.259 | -7.791 | 16.165 |
| ATOM | 316 | CD | GLU | A | 34 | -8.763 | -8.552 | 17.404 |
| ATOM | 317 | OE1 | GLU | A | 34 | -7.670 | -9.193 | 17.368 |
| ATOM | 318 | OE2 | GLU | A | 34 | -9.482 | -8.497 | 18.407 |
| ATOM | 319 | N | GLU | A | 35 | -10.152 | -7.480 | 12.568 |
| ATOM | 320 | H | GLU | A | 35 | -9.485 | -8.208 | 12.407 |
| ATOM | 321 | CA | GLU | A | 35 | -11.352 | -7.466 | 11.773 |
| ATOM | 322 | C | GLU | A | 35 | -12.631 | -7.520 | 12.571 |
| ATOM | 323 | O | GLU | A | 35 | -12.814 | -8.294 | 13.528 |
| ATOM | 324 | CB | GLU | A | 35 | -11.237 | -8.536 | 10.707 |
| ATOM | 325 | CG | GLU | A | 35 | -9.945 | -8.280 | 9.907 |
| ATOM | 326 | CD | GLU | A | 35 | -9.872 | -8.872 | 8.486 |
| ATOM | 327 | OE1 | GLU | A | 35 | -10.612 | -8.401 | 7.603 |
| ATOM | 328 | OE2 | GLU | A | 35 | -9.024 | -9.776 | 8.261 |
| ATOM | 329 | N | MET | A | 36 | -13.580 | -6.598 | 12.278 |
| ATOM | 330 | H | MET | A | 36 | -13.439 | -5.967 | 11.515 |
| ATOM | 331 | CA | MET | A | 36 | -14.819 | -6.495 | 13.052 |
| ATOM | 332 | C | MET | A | 36 | -15.826 | -5.635 | 12.271 |
| ATOM | 333 | O | MET | A | 36 | -15.514 | -4.828 | 11.371 |
| ATOM | 334 | CB | MET | A | 36 | -14.593 | -5.845 | 14.428 |
| ATOM | 335 | CG | MET | A | 36 | -14.279 | -4.353 | 14.417 |

FIG. 1F

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 336 | SD | MET | A | 36 | -14.251 | -3.718 | 16.099 |
| ATOM | 337 | CE | MET | A | 36 | -12.487 | -3.846 | 16.409 |
| ATOM | 338 | N | SER | A | 37 | -17.130 | -5.776 | 12.590 |
| ATOM | 339 | H | SER | A | 37 | -17.399 | -6.431 | 13.296 |
| ATOM | 340 | CA | SER | A | 37 | -18.155 | -5.005 | 11.940 |
| ATOM | 341 | C | SER | A | 37 | -18.286 | -3.693 | 12.657 |
| ATOM | 342 | O | SER | A | 37 | -18.593 | -3.624 | 13.865 |
| ATOM | 343 | CB | SER | A | 37 | -19.506 | -5.688 | 12.032 |
| ATOM | 344 | OG | SER | A | 37 | -19.455 | -7.054 | 11.716 |
| ATOM | 345 | HG | SER | A | 37 | -20.367 | -7.457 | 11.791 |
| ATOM | 346 | N | LEU | A | 38 | -18.185 | -2.569 | 11.933 |
| ATOM | 347 | H | LEU | A | 38 | -17.956 | -2.625 | 10.952 |
| ATOM | 348 | CA | LEU | A | 38 | -18.557 | -1.247 | 12.465 |
| ATOM | 349 | C | LEU | A | 38 | -19.630 | -0.605 | 11.572 |
| ATOM | 350 | O | LEU | A | 38 | -19.706 | -0.939 | 10.391 |
| ATOM | 351 | CB | LEU | A | 38 | -17.315 | -0.346 | 12.588 |
| ATOM | 352 | CG | LEU | A | 38 | -16.246 | -0.818 | 13.596 |
| ATOM | 353 | CD1 | LEU | A | 38 | -14.998 | 0.073 | 13.489 |
| ATOM | 354 | CD2 | LEU | A | 38 | -16.756 | -0.787 | 15.046 |
| ATOM | 355 | N | PRO | A | 39 | -20.455 | 0.321 | 12.108 |
| ATOM | 356 | CA | PRO | A | 39 | -21.460 | 1.053 | 11.339 |
| ATOM | 357 | C | PRO | A | 39 | -20.824 | 2.176 | 10.502 |
| ATOM | 358 | O | PRO | A | 39 | -19.654 | 2.519 | 10.685 |
| ATOM | 359 | CB | PRO | A | 39 | -22.430 | 1.607 | 12.389 |
| ATOM | 360 | CG | PRO | A | 39 | -21.531 | 1.845 | 13.600 |
| ATOM | 361 | CD | PRO | A | 39 | -20.539 | 0.686 | 13.517 |
| ATOM | 362 | N | GLY | A | 40 | -21.620 | 2.749 | 9.586 |
| ATOM | 363 | H | GLY | A | 40 | -22.569 | 2.417 | 9.493 |
| ATOM | 364 | CA | GLY | A | 40 | -21.203 | 3.811 | 8.678 |
| ATOM | 365 | C | GLY | A | 40 | -20.836 | 3.262 | 7.298 |
| ATOM | 366 | O | GLY | A | 40 | -21.405 | 2.268 | 6.845 |
| ATOM | 367 | N | LYS | A | 41 | -19.895 | 3.945 | 6.631 |
| ATOM | 368 | H | LYS | A | 41 | -19.496 | 4.761 | 7.071 |
| ATOM | 369 | CA | LYS | A | 41 | -19.323 | 3.558 | 5.343 |
| ATOM | 370 | C | LYS | A | 41 | -17.798 | 3.757 | 5.371 |
| ATOM | 371 | O | LYS | A | 41 | -17.263 | 4.462 | 6.229 |
| ATOM | 372 | CB | LYS | A | 41 | -20.025 | 4.352 | 4.224 |
| ATOM | 373 | CG | LYS | A | 41 | -19.703 | 3.839 | 2.810 |
| ATOM | 374 | CD | LYS | A | 41 | -20.610 | 4.486 | 1.757 |
| ATOM | 375 | CE | LYS | A | 41 | -20.240 | 3.964 | 0.366 |
| ATOM | 376 | NZ | LYS | A | 41 | -21.097 | 4.552 | -0.678 |
| ATOM | 377 | 1HZ | LYS | A | 41 | -20.824 | 4.189 | -1.580 |
| ATOM | 378 | 3HZ | LYS | A | 41 | -20.993 | 5.556 | -0.673 |
| ATOM | 379 | 2HZ | LYS | A | 41 | -22.061 | 4.311 | -0.498 |
| ATOM | 380 | N | TRP | A | 42 | -17.104 | 3.091 | 4.439 |
| ATOM | 381 | H | TRP | A | 42 | -17.620 | 2.548 | 3.762 |
| ATOM | 382 | CA | TRP | A | 42 | -15.654 | 2.932 | 4.423 |
| ATOM | 383 | C | TRP | A | 42 | -15.105 | 2.852 | 2.994 |
| ATOM | 384 | O | TRP | A | 42 | -15.845 | 2.702 | 2.021 |
| ATOM | 385 | CB | TRP | A | 42 | -15.279 | 1.675 | 5.236 |
| ATOM | 386 | CG | TRP | A | 42 | -16.214 | 0.514 | 5.094 |
| ATOM | 387 | CD1 | TRP | A | 42 | -16.230 | -0.402 | 4.101 |
| ATOM | 388 | CD2 | TRP | A | 42 | -17.355 | 0.203 | 5.942 |
| ATOM | 389 | NE1 | TRP | A | 42 | -17.297 | -1.260 | 4.281 |
| ATOM | 390 | HE1 | TRP | A | 42 | -17.504 | -2.015 | 3.644 |
| ATOM | 391 | CE2 | TRP | A | 42 | -18.045 | -0.914 | 5.389 |

FIG. 11G

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 392 | CE3 | TRP | A | 42 | -17.896 | 0.792 | 7.103 |
| ATOM | 393 | CZ2 | TRP | A | 42 | -19.224 | -1.421 | 5.959 |
| ATOM | 394 | CZ3 | TRP | A | 42 | -19.077 | 0.298 | 7.675 |
| ATOM | 395 | CH2 | TRP | A | 42 | -19.741 | -0.806 | 7.112 |
| ATOM | 396 | N | LYS | A | 43 | -13.771 | 2.932 | 2.911 |
| ATOM | 397 | H | LYS | A | 43 | -13.260 | 3.058 | 3.773 |
| ATOM | 398 | CA | LYS | A | 43 | -12.951 | 2.802 | 1.713 |
| ATOM | 399 | C | LYS | A | 43 | -11.773 | 1.859 | 2.012 |
| ATOM | 400 | O | LYS | A | 43 | -11.359 | 1.760 | 3.166 |
| ATOM | 401 | CB | LYS | A | 43 | -12.451 | 4.193 | 1.270 |
| ATOM | 402 | CG | LYS | A | 43 | -11.724 | 4.979 | 2.383 |
| ATOM | 403 | CD | LYS | A | 43 | -11.060 | 6.267 | 1.873 |
| ATOM | 404 | CE | LYS | A | 43 | -9.784 | 6.001 | 1.065 |
| ATOM | 405 | NZ | LYS | A | 43 | -8.700 | 5.458 | 1.903 |
| ATOM | 406 | 1HZ | LYS | A | 43 | -7.876 | 5.315 | 1.338 |
| ATOM | 407 | 3HZ | LYS | A | 43 | -8.993 | 4.576 | 2.300 |
| ATOM | 408 | 2HZ | LYS | A | 43 | -8.493 | 6.108 | 2.647 |
| ATOM | 409 | N | PRO | A | 44 | -11.177 | 1.197 | 1.004 |
| ATOM | 410 | CA | PRO | A | 44 | -9.947 | 0.435 | 1.187 |
| ATOM | 411 | C | PRO | A | 44 | -8.760 | 1.392 | 1.379 |
| ATOM | 412 | O | PRO | A | 44 | -8.711 | 2.434 | 0.720 |
| ATOM | 413 | CB | PRO | A | 44 | -9.808 | -0.393 | -0.095 |
| ATOM | 414 | CG | PRO | A | 44 | -10.501 | 0.458 | -1.159 |
| ATOM | 415 | CD | PRO | A | 44 | -11.630 | 1.132 | -0.380 |
| ATOM | 416 | N | LYS | A | 45 | -7.790 | 1.030 | 2.240 |
| ATOM | 417 | H | LYS | A | 45 | -7.912 | 0.227 | 2.824 |
| ATOM | 418 | CA | LYS | A | 45 | -6.547 | 1.747 | 2.314 |
| ATOM | 419 | C | LYS | A | 45 | -5.493 | 0.683 | 2.507 |
| ATOM | 420 | O | LYS | A | 45 | -5.780 | -0.470 | 2.869 |
| ATOM | 421 | CB | LYS | A | 45 | -6.594 | 2.699 | 3.524 |
| ATOM | 422 | CG | LYS | A | 45 | -5.463 | 3.744 | 3.609 |
| ATOM | 423 | CD | LYS | A | 45 | -5.340 | 4.289 | 5.052 |
| ATOM | 424 | CE | LYS | A | 45 | -4.262 | 5.383 | 5.204 |
| ATOM | 425 | NZ | LYS | A | 45 | -2.907 | 4.911 | 4.916 |
| ATOM | 426 | 1HZ | LYS | A | 45 | -2.260 | 5.664 | 5.032 |
| ATOM | 427 | 3HZ | LYS | A | 45 | -2.864 | 4.577 | 3.975 |
| ATOM | 428 | 2HZ | LYS | A | 45 | -2.672 | 4.169 | 5.544 |
| ATOM | 429 | N | MET | A | 46 | -4.224 | 0.949 | 2.193 |
| ATOM | 430 | H | MET | A | 46 | -3.998 | 1.805 | 1.728 |
| ATOM | 431 | CA | MET | A | 46 | -3.157 | 0.027 | 2.509 |
| ATOM | 432 | C | MET | A | 46 | -2.417 | 0.701 | 3.627 |
| ATOM | 433 | O | MET | A | 46 | -2.259 | 1.937 | 3.634 |
| ATOM | 434 | CB | MET | A | 46 | -2.166 | -0.088 | 1.379 |
| ATOM | 435 | CG | MET | A | 46 | -2.782 | -0.366 | 0.053 |
| ATOM | 436 | SD | MET | A | 46 | -3.076 | -2.108 | -0.118 |
| ATOM | 437 | CE | MET | A | 46 | -1.417 | -2.652 | -0.186 |
| ATOM | 438 | N | ILE | A | 47 | -1.827 | -0.016 | 4.586 |
| ATOM | 439 | H | ILE | A | 47 | -2.010 | -0.997 | 4.655 |
| ATOM | 440 | CA | ILE | A | 47 | -0.922 | 0.586 | 5.539 |
| ATOM | 441 | C | ILE | A | 47 | 0.233 | -0.372 | 5.654 |
| ATOM | 442 | O | ILE | A | 47 | 0.135 | -1.584 | 5.356 |
| ATOM | 443 | CB | ILE | A | 47 | -1.550 | 0.836 | 6.923 |
| ATOM | 444 | CG1 | ILE | A | 47 | -2.459 | -0.301 | 7.354 |
| ATOM | 445 | CG2 | ILE | A | 47 | -2.248 | 2.164 | 6.995 |
| ATOM | 446 | CD1 | ILE | A | 47 | -1.724 | -1.336 | 8.111 |
| ATOM | 447 | N | GLY | A | 48 | 1.420 | 0.089 | 6.043 |

FIG. 11H

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 448 | H | GLY | A | 48 | 1.509 | 1.040 | 6.339 |
| ATOM | 449 | CA | GLY | A | 48 | 2.584 | -0.753 | 6.048 |
| ATOM | 450 | C | GLY | A | 48 | 3.280 | -0.657 | 7.376 |
| ATOM | 451 | O | GLY | A | 48 | 3.050 | 0.190 | 8.265 |
| ATOM | 452 | N | GLY | A | 49 | 4.197 | -1.617 | 7.603 |
| ATOM | 453 | H | GLY | A | 49 | 4.375 | -2.308 | 6.902 |
| ATOM | 454 | CA | GLY | A | 49 | 4.936 | -1.684 | 8.828 |
| ATOM | 455 | C | GLY | A | 49 | 6.105 | -2.589 | 8.533 |
| ATOM | 456 | O | GLY | A | 49 | 6.482 | -2.807 | 7.370 |
| ATOM | 457 | N | ILE | A | 50 | 6.761 | -3.173 | 9.552 |
| ATOM | 458 | H | ILE | A | 50 | 6.552 | -2.908 | 10.493 |
| ATOM | 459 | CA | ILE | A | 50 | 7.772 | -4.184 | 9.344 |
| ATOM | 460 | C | ILE | A | 50 | 7.148 | -5.317 | 8.566 |
| ATOM | 461 | O | ILE | A | 50 | 5.981 | -5.734 | 8.772 |
| ATOM | 462 | CB | ILE | A | 50 | 8.258 | -4.686 | 10.722 |
| ATOM | 463 | CG1 | ILE | A | 50 | 9.257 | -3.714 | 11.382 |
| ATOM | 464 | CG2 | ILE | A | 50 | 8.813 | -6.134 | 10.693 |
| ATOM | 465 | CD1 | ILE | A | 50 | 10.580 | -3.498 | 10.628 |
| ATOM | 466 | N | GLY | A | 51 | 7.847 | -5.891 | 7.596 |
| ATOM | 467 | H | GLY | A | 51 | 8.772 | -5.569 | 7.395 |
| ATOM | 468 | CA | GLY | A | 51 | 7.265 | -6.966 | 6.850 |
| ATOM | 469 | C | GLY | A | 51 | 6.519 | -6.559 | 5.591 |
| ATOM | 470 | O | GLY | A | 51 | 6.430 | -7.318 | 4.634 |
| ATOM | 471 | N | GLY | A | 52 | 5.886 | -5.375 | 5.517 |
| ATOM | 472 | H | GLY | A | 52 | 5.990 | -4.710 | 6.257 |
| ATOM | 473 | CA | GLY | A | 52 | 5.108 | -5.227 | 4.320 |
| ATOM | 474 | C | GLY | A | 52 | 3.832 | -4.415 | 4.516 |
| ATOM | 475 | O | GLY | A | 52 | 3.654 | -3.624 | 5.467 |
| ATOM | 476 | N | PHE | A | 53 | 2.886 | -4.518 | 3.559 |
| ATOM | 477 | H | PHE | A | 53 | 3.013 | -5.161 | 2.804 |
| ATOM | 478 | CA | PHE | A | 53 | 1.653 | -3.720 | 3.566 |
| ATOM | 479 | C | PHE | A | 53 | 0.494 | -4.651 | 3.783 |
| ATOM | 480 | O | PHE | A | 53 | 0.448 | -5.816 | 3.336 |
| ATOM | 481 | CB | PHE | A | 53 | 1.424 | -3.022 | 2.221 |
| ATOM | 482 | CG | PHE | A | 53 | 2.363 | -1.896 | 2.008 |
| ATOM | 483 | CD1 | PHE | A | 53 | 3.615 | -2.135 | 1.447 |
| ATOM | 484 | CD2 | PHE | A | 53 | 2.011 | -0.608 | 2.414 |
| ATOM | 485 | CE1 | PHE | A | 53 | 4.514 | -1.087 | 1.275 |
| ATOM | 486 | CE2 | PHE | A | 53 | 2.925 | 0.446 | 2.237 |
| ATOM | 487 | CZ | PHE | A | 53 | 4.172 | 0.202 | 1.668 |
| ATOM | 488 | N | ILE | A | 54 | -0.554 | -4.173 | 4.439 |
| ATOM | 489 | H | ILE | A | 54 | -0.491 | -3.285 | 4.895 |
| ATOM | 490 | CA | ILE | A | 54 | -1.789 | -4.911 | 4.509 |
| ATOM | 491 | C | ILE | A | 54 | -2.903 | -3.995 | 4.033 |
| ATOM | 492 | O | ILE | A | 54 | -2.751 | -2.770 | 3.855 |
| ATOM | 493 | CB | ILE | A | 54 | -2.034 | -5.535 | 5.904 |
| ATOM | 494 | CG1 | ILE | A | 54 | -2.343 | -4.481 | 6.988 |
| ATOM | 495 | CG2 | ILE | A | 54 | -0.799 | -6.318 | 6.314 |
| ATOM | 496 | CD1 | ILE | A | 54 | -3.010 | -5.089 | 8.246 |
| ATOM | 497 | N | LYS | A | 55 | -4.029 | -4.577 | 3.560 |
| ATOM | 498 | H | LYS | A | 55 | -4.084 | -5.574 | 3.501 |
| ATOM | 499 | CA | LYS | A | 55 | -5.177 | -3.798 | 3.129 |
| ATOM | 500 | C | LYS | A | 55 | -6.115 | -3.726 | 4.300 |
| ATOM | 501 | O | LYS | A | 55 | -6.422 | -4.707 | 5.023 |
| ATOM | 502 | CB | LYS | A | 55 | -5.928 | -4.461 | 1.938 |
| ATOM | 503 | CG | LYS | A | 55 | -6.853 | -3.547 | 1.106 |

FIG. 1 II

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 504 | CD | LYS | A | 55 | -8.267 | -3.332 | 1.714 |
| ATOM | 505 | CE | LYS | A | 55 | -9.303 | -4.392 | 1.301 |
| ATOM | 506 | NZ | LYS | A | 55 | -10.521 | -4.453 | 2.192 |
| ATOM | 507 | 1HZ | LYS | A | 55 | -11.142 | -5.162 | 1.859 |
| ATOM | 508 | 3HZ | LYS | A | 55 | -10.987 | -3.569 | 2.180 |
| ATOM | 509 | 2HZ | LYS | A | 55 | -10.240 | -4.669 | 3.127 |
| ATOM | 510 | N | VAL | A | 56 | -6.599 | -2.509 | 4.619 |
| ATOM | 511 | H | VAL | A | 56 | -6.337 | -1.713 | 4.073 |
| ATOM | 512 | CA | VAL | A | 56 | -7.494 | -2.311 | 5.735 |
| ATOM | 513 | C | VAL | A | 56 | -8.711 | -1.584 | 5.236 |
| ATOM | 514 | O | VAL | A | 56 | -8.767 | -1.029 | 4.114 |
| ATOM | 515 | CB | VAL | A | 56 | -6.759 | -1.475 | 6.812 |
| ATOM | 516 | CG1 | VAL | A | 56 | -5.569 | -2.209 | 7.385 |
| ATOM | 517 | CG2 | VAL | A | 56 | -6.287 | -0.108 | 6.268 |
| ATOM | 518 | N | ARG | A | 57 | -9.784 | -1.539 | 6.005 |
| ATOM | 519 | H | ARG | A | 57 | -9.835 | -2.117 | 6.819 |
| ATOM | 520 | CA | ARG | A | 57 | -10.855 | -0.648 | 5.638 |
| ATOM | 521 | C | ARG | A | 57 | -10.738 | 0.534 | 6.554 |
| ATOM | 522 | O | ARG | A | 57 | -10.558 | 0.449 | 7.789 |
| ATOM | 523 | CB | ARG | A | 57 | -12.219 | -1.271 | 5.835 |
| ATOM | 524 | CG | ARG | A | 57 | -12.480 | -2.452 | 4.952 |
| ATOM | 525 | CD | ARG | A | 57 | -13.834 | -3.051 | 5.195 |
| ATOM | 526 | NE | ARG | A | 57 | -14.122 | -4.137 | 4.270 |
| ATOM | 527 | HE | ARG | A | 57 | -13.442 | -4.347 | 3.568 |
| ATOM | 528 | CZ | ARG | A | 57 | -15.243 | -4.851 | 4.324 |
| ATOM | 529 | NH1 | ARG | A | 57 | -16.175 | -4.624 | 5.243 |
| ATOM | 530 | 2HH1 | ARG | A | 57 | -16.044 | -3.899 | 5.920 |
| ATOM | 531 | 1HH1 | ARG | A | 57 | -17.008 | -5.178 | 5.258 |
| ATOM | 532 | NH2 | ARG | A | 57 | -15.433 | -5.822 | 3.434 |
| ATOM | 533 | 1HH2 | ARG | A | 57 | -16.270 | -6.368 | 3.461 |
| ATOM | 534 | 2HH2 | ARG | A | 57 | -14.738 | -6.006 | 2.738 |
| ATOM | 535 | N | GLN | A | 58 | -10.881 | 1.741 | 6.036 |
| ATOM | 536 | H | GLN | A | 58 | -11.030 | 1.844 | 5.053 |
| ATOM | 537 | CA | GLN | A | 58 | -10.830 | 2.922 | 6.839 |
| ATOM | 538 | C | GLN | A | 58 | -12.231 | 3.342 | 7.205 |
| ATOM | 539 | O | GLN | A | 58 | -13.106 | 3.608 | 6.359 |
| ATOM | 540 | CB | GLN | A | 58 | -10.208 | 4.038 | 6.030 |
| ATOM | 541 | CG | GLN | A | 58 | -10.055 | 5.293 | 6.817 |
| ATOM | 542 | CD | GLN | A | 58 | -9.632 | 6.411 | 5.927 |
| ATOM | 543 | OE1 | GLN | A | 58 | -10.379 | 7.334 | 5.662 |
| ATOM | 544 | NE2 | GLN | A | 58 | -8.412 | 6.303 | 5.437 |
| ATOM | 545 | 1HE2 | GLN | A | 58 | -8.047 | 7.009 | 4.830 |
| ATOM | 546 | 2HE2 | GLN | A | 58 | -7.843 | 5.514 | 5.668 |
| ATOM | 547 | N | TYR | A | 59 | -12.527 | 3.516 | 8.509 |
| ATOM | 548 | H | TYR | A | 59 | -11.877 | 3.219 | 9.209 |
| ATOM | 549 | CA | TYR | A | 59 | -13.769 | 4.125 | 8.933 |
| ATOM | 550 | C | TYR | A | 59 | -13.411 | 5.452 | 9.565 |
| ATOM | 551 | O | TYR | A | 59 | -12.416 | 5.592 | 10.310 |
| ATOM | 552 | CB | TYR | A | 59 | -14.517 | 3.252 | 9.957 |
| ATOM | 553 | CG | TYR | A | 59 | -14.287 | 1.770 | 9.723 |
| ATOM | 554 | CD1 | TYR | A | 59 | -13.007 | 1.269 | 9.457 |
| ATOM | 555 | CD2 | TYR | A | 59 | -15.346 | 0.865 | 9.766 |
| ATOM | 556 | CE1 | TYR | A | 59 | -12.797 | -0.092 | 9.240 |
| ATOM | 557 | CE2 | TYR | A | 59 | -15.148 | -0.494 | 9.551 |
| ATOM | 558 | CZ | TYR | A | 59 | -13.873 | -0.972 | 9.287 |
| ATOM | 559 | OH | TYR | A | 59 | -13.721 | -2.311 | 9.079 |

FIG. 11J

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 560 | HH | TYR | A | 59 | -14.606 | -2.771 | 9.154 |
| ATOM | 561 | N | ASP | A | 60 | -14.151 | 6.542 | 9.300 |
| ATOM | 562 | H | ASP | A | 60 | -14.954 | 6.464 | 8.709 |
| ATOM | 563 | CA | ASP | A | 60 | -13.822 | 7.836 | 9.846 |
| ATOM | 564 | C | ASP | A | 60 | -14.782 | 8.226 | 10.947 |
| ATOM | 565 | O | ASP | A | 60 | -15.941 | 7.765 | 11.053 |
| ATOM | 566 | CB | ASP | A | 60 | -13.861 | 8.942 | 8.769 |
| ATOM | 567 | CG | ASP | A | 60 | -12.735 | 8.830 | 7.725 |
| ATOM | 568 | OD1 | ASP | A | 60 | -11.545 | 8.874 | 8.075 |
| ATOM | 569 | OD2 | ASP | A | 60 | -13.060 | 8.702 | 6.544 |
| ATOM | 570 | N | GLN | A | 61 | -14.339 | 9.154 | 11.833 |
| ATOM | 571 | H | GLN | A | 61 | -13.385 | 9.451 | 11.804 |
| ATOM | 572 | CA | GLN | A | 61 | -15.151 | 9.804 | 12.885 |
| ATOM | 573 | C | GLN | A | 61 | -15.839 | 8.803 | 13.802 |
| ATOM | 574 | O | GLN | A | 61 | -17.008 | 8.893 | 14.229 |
| ATOM | 575 | CB | GLN | A | 61 | -16.097 | 10.908 | 12.338 |
| ATOM | 576 | CG | GLN | A | 61 | -16.239 | 12.133 | 13.262 |
| ATOM | 577 | CD | GLN | A | 61 | -16.910 | 13.366 | 12.629 |
| ATOM | 578 | OE1 | GLN | A | 61 | -16.509 | 13.854 | 11.586 |
| ATOM | 579 | NE2 | GLN | A | 61 | -17.937 | 13.887 | 13.292 |
| ATOM | 580 | 1HE2 | GLN | A | 61 | -18.416 | 14.689 | 12.934 |
| ATOM | 581 | 2HE2 | GLN | A | 61 | -18.239 | 13.482 | 14.155 |
| ATOM | 582 | N | ILE | A | 62 | -15.060 | 7.760 | 14.175 |
| ATOM | 583 | H | ILE | A | 62 | -14.111 | 7.714 | 13.862 |
| ATOM | 584 | CA | ILE | A | 62 | -15.557 | 6.705 | 15.015 |
| ATOM | 585 | C | ILE | A | 62 | -15.251 | 7.057 | 16.447 |
| ATOM | 586 | O | ILE | A | 62 | -14.198 | 7.613 | 16.837 |
| ATOM | 587 | CB | ILE | A | 62 | -14.829 | 5.397 | 14.653 |
| ATOM | 588 | CG1 | ILE | A | 62 | -15.253 | 4.966 | 13.258 |
| ATOM | 589 | CG2 | ILE | A | 62 | -15.106 | 4.271 | 15.675 |
| ATOM | 590 | CD1 | ILE | A | 62 | -16.779 | 4.788 | 13.116 |
| ATOM | 591 | N | LEU | A | 63 | -16.242 | 6.807 | 17.320 |
| ATOM | 592 | H | LEU | A | 63 | -17.089 | 6.383 | 17.000 |
| ATOM | 593 | CA | LEU | A | 63 | -16.127 | 7.131 | 18.719 |
| ATOM | 594 | C | LEU | A | 63 | -15.518 | 5.942 | 19.425 |
| ATOM | 595 | O | LEU | A | 63 | -15.869 | 4.753 | 19.269 |
| ATOM | 596 | CB | LEU | A | 63 | -17.512 | 7.428 | 19.282 |
| ATOM | 597 | CG | LEU | A | 63 | -17.660 | 7.598 | 20.813 |
| ATOM | 598 | CD1 | LEU | A | 63 | -16.711 | 8.632 | 21.404 |
| ATOM | 599 | CD2 | LEU | A | 63 | -19.089 | 7.963 | 21.201 |
| ATOM | 600 | N | ILE | A | 64 | -14.511 | 6.211 | 20.219 |
| ATOM | 601 | H | ILE | A | 64 | -14.185 | 7.153 | 20.305 |
| ATOM | 602 | CA | ILE | A | 64 | -13.862 | 5.178 | 20.972 |
| ATOM | 603 | C | ILE | A | 64 | -13.529 | 5.744 | 22.325 |
| ATOM | 604 | O | ILE | A | 64 | -13.396 | 6.959 | 22.602 |
| ATOM | 605 | CB | ILE | A | 64 | -12.618 | 4.716 | 20.231 |
| ATOM | 606 | CG1 | ILE | A | 64 | -11.925 | 3.573 | 20.949 |
| ATOM | 607 | CG2 | ILE | A | 64 | -11.690 | 5.865 | 19.950 |
| ATOM | 608 | CD1 | ILE | A | 64 | -10.905 | 2.888 | 20.062 |
| ATOM | 609 | N | GLU | A | 65 | -13.396 | 4.815 | 23.294 |
| ATOM | 610 | H | GLU | A | 65 | -13.443 | 3.844 | 23.059 |
| ATOM | 611 | CA | GLU | A | 65 | -13.186 | 5.174 | 24.670 |
| ATOM | 612 | C | GLU | A | 65 | -12.024 | 4.360 | 25.165 |
| ATOM | 613 | O | GLU | A | 65 | -11.943 | 3.112 | 25.056 |
| ATOM | 614 | CB | GLU | A | 65 | -14.459 | 4.823 | 25.405 |
| ATOM | 615 | CG | GLU | A | 65 | -14.739 | 5.610 | 26.646 |

FIG. 11K

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 616 | CD | GLU | A | 65 | -16.131 | 5.353 | 27.115 |
| ATOM | 617 | OE1 | GLU | A | 65 | -17.090 | 5.785 | 26.413 |
| ATOM | 618 | OE2 | GLU | A | 65 | -16.269 | 4.708 | 28.163 |
| ATOM | 619 | N | ILE | A | 66 | -10.971 | 5.008 | 25.610 |
| ATOM | 620 | H | ILE | A | 66 | -11.009 | 6.002 | 25.717 |
| ATOM | 621 | CA | ILE | A | 66 | -9.762 | 4.317 | 25.947 |
| ATOM | 622 | C | ILE | A | 66 | -9.571 | 4.586 | 27.413 |
| ATOM | 623 | O | ILE | A | 66 | -9.422 | 5.732 | 27.880 |
| ATOM | 624 | CB | ILE | A | 66 | -8.600 | 4.907 | 25.126 |
| ATOM | 625 | CG1 | ILE | A | 66 | -8.838 | 4.669 | 23.633 |
| ATOM | 626 | CG2 | ILE | A | 66 | -7.231 | 4.326 | 25.554 |
| ATOM | 627 | CD1 | ILE | A | 66 | -8.951 | 5.982 | 22.856 |
| ATOM | 628 | N | CYS | A | 67 | -9.776 | 3.567 | 28.261 |
| ATOM | 629 | H | CYS | A | 67 | -9.989 | 2.659 | 27.902 |
| ATOM | 630 | CA | CYS | A | 67 | -9.698 | 3.740 | 29.687 |
| ATOM | 631 | C | CYS | A | 67 | -10.673 | 4.871 | 30.088 |
| ATOM | 632 | O | CYS | A | 67 | -10.393 | 5.716 | 30.958 |
| ATOM | 633 | CB | CYS | A | 67 | -8.251 | 4.003 | 30.156 |
| ATOM | 634 | SG | CYS | A | 67 | -7.170 | 2.529 | 30.217 |
| ATOM | 635 | N | GLY | A | 68 | -11.877 | 4.947 | 29.499 |
| ATOM | 636 | H | GLY | A | 68 | -12.125 | 4.286 | 28.791 |
| ATOM | 637 | CA | GLY | A | 68 | -12.788 | 5.984 | 29.903 |
| ATOM | 638 | C | GLY | A | 68 | -12.581 | 7.322 | 29.241 |
| ATOM | 639 | O | GLY | A | 68 | -13.404 | 8.253 | 29.376 |
| ATOM | 640 | N | HIS | A | 69 | -11.504 | 7.545 | 28.471 |
| ATOM | 641 | H | HIS | A | 69 | -10.817 | 6.827 | 28.360 |
| ATOM | 642 | CA | HIS | A | 69 | -11.305 | 8.800 | 27.793 |
| ATOM | 643 | C | HIS | A | 69 | -11.838 | 8.679 | 26.399 |
| ATOM | 644 | O | HIS | A | 69 | -11.516 | 7.742 | 25.630 |
| ATOM | 645 | CB | HIS | A | 69 | -9.831 | 9.128 | 27.724 |
| ATOM | 646 | CG | HIS | A | 69 | -9.276 | 9.286 | 29.081 |
| ATOM | 647 | ND1 | HIS | A | 69 | -9.317 | 10.484 | 29.778 |
| ATOM | 648 | HD1 | HIS | A | 69 | -9.688 | 11.347 | 29.436 |
| ATOM | 649 | CD2 | HIS | A | 69 | -8.723 | 8.352 | 29.912 |
| ATOM | 650 | CE1 | HIS | A | 69 | -8.783 | 10.254 | 30.947 |
| ATOM | 651 | NE2 | HIS | A | 69 | -8.405 | 8.990 | 31.091 |
| ATOM | 652 | N | LYS | A | 70 | -12.768 | 9.561 | 25.973 |
| ATOM | 653 | H | LYS | A | 70 | -13.084 | 10.284 | 26.588 |
| ATOM | 654 | CA | LYS | A | 70 | -13.325 | 9.492 | 24.646 |
| ATOM | 655 | C | LYS | A | 70 | -12.346 | 10.074 | 23.653 |
| ATOM | 656 | O | LYS | A | 70 | -11.587 | 11.055 | 23.864 |
| ATOM | 657 | CB | LYS | A | 70 | -14.645 | 10.285 | 24.536 |
| ATOM | 658 | CG | LYS | A | 70 | -15.837 | 9.703 | 25.330 |
| ATOM | 659 | CD | LYS | A | 70 | -17.105 | 10.593 | 25.286 |
| ATOM | 660 | CE | LYS | A | 70 | -18.293 | 10.011 | 26.092 |
| ATOM | 661 | NZ | LYS | A | 70 | -18.802 | 8.702 | 25.608 |
| ATOM | 662 | 1HZ | LYS | A | 70 | -19.563 | 8.406 | 26.185 |
| ATOM | 663 | 3HZ | LYS | A | 70 | -18.069 | 8.023 | 25.650 |
| ATOM | 664 | 2HZ | LYS | A | 70 | -19.116 | 8.795 | 24.663 |
| ATOM | 665 | N | ALA | A | 71 | -12.323 | 9.485 | 22.446 |
| ATOM | 666 | H | ALA | A | 71 | -12.813 | 8.625 | 22.305 |
| ATOM | 667 | CA | ALA | A | 71 | -11.616 | 10.044 | 21.333 |
| ATOM | 668 | C | ALA | A | 71 | -12.529 | 9.795 | 20.171 |
| ATOM | 669 | O | ALA | A | 71 | -13.351 | 8.850 | 20.146 |
| ATOM | 670 | CB | ALA | A | 71 | -10.292 | 9.358 | 21.143 |
| ATOM | 671 | N | ILE | A | 72 | -12.559 | 10.685 | 19.149 |

FIG. 1 IL

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 672 | H | ILE | A | 72 | -12.006 | 11.517 | 19.200 |
| ATOM | 673 | CA | ILE | A | 72 | -13.376 | 10.474 | 17.963 |
| ATOM | 674 | C | ILE | A | 72 | -12.480 | 10.662 | 16.771 |
| ATOM | 675 | O | ILE | A | 72 | -11.858 | 11.720 | 16.550 |
| ATOM | 676 | CB | ILE | A | 72 | -14.541 | 11.464 | 17.882 |
| ATOM | 677 | CG1 | ILE | A | 72 | -15.306 | 11.455 | 19.196 |
| ATOM | 678 | CG2 | ILE | A | 72 | -15.429 | 11.203 | 16.651 |
| ATOM | 679 | CD1 | ILE | A | 72 | -16.446 | 12.415 | 19.176 |
| ATOM | 680 | N | GLY | A | 73 | -12.252 | 9.633 | 15.958 |
| ATOM | 681 | H | GLY | A | 73 | -12.778 | 8.789 | 16.067 |
| ATOM | 682 | CA | GLY | A | 73 | -11.253 | 9.755 | 14.938 |
| ATOM | 683 | C | GLY | A | 73 | -11.283 | 8.554 | 14.034 |
| ATOM | 684 | O | GLY | A | 73 | -12.211 | 7.706 | 14.006 |
| ATOM | 685 | N | THR | A | 74 | -10.247 | 8.428 | 13.182 |
| ATOM | 686 | H | THR | A | 74 | -9.471 | 9.055 | 13.250 |
| ATOM | 687 | CA | THR | A | 74 | -10.201 | 7.416 | 12.158 |
| ATOM | 688 | C | THR | A | 74 | -9.674 | 6.134 | 12.760 |
| ATOM | 689 | O | THR | A | 74 | -8.670 | 6.034 | 13.497 |
| ATOM | 690 | CB | THR | A | 74 | -9.298 | 7.895 | 11.048 |
| ATOM | 691 | OG1 | THR | A | 74 | -9.910 | 9.019 | 10.441 |
| ATOM | 692 | HG1 | THR | A | 74 | -9.335 | 9.362 | 9.698 |
| ATOM | 693 | CG2 | THR | A | 74 | -9.088 | 6.823 | 9.946 |
| ATOM | 694 | N | VAL | A | 75 | -10.318 | 5.027 | 12.327 |
| ATOM | 695 | H | VAL | A | 75 | -11.066 | 5.114 | 11.669 |
| ATOM | 696 | CA | VAL | A | 75 | -9.968 | 3.717 | 12.778 |
| ATOM | 697 | C | VAL | A | 75 | -9.906 | 2.843 | 11.551 |
| ATOM | 698 | O | VAL | A | 75 | -10.803 | 2.807 | 10.681 |
| ATOM | 699 | CB | VAL | A | 75 | -11.044 | 3.250 | 13.737 |
| ATOM | 700 | CG1 | VAL | A | 75 | -11.021 | 1.721 | 13.943 |
| ATOM | 701 | CG2 | VAL | A | 75 | -10.915 | 4.019 | 15.034 |
| ATOM | 702 | N | LEU | A | 76 | -8.768 | 2.139 | 11.366 |
| ATOM | 703 | H | LEU | A | 76 | -8.002 | 2.260 | 11.998 |
| ATOM | 704 | CA | LEU | A | 76 | -8.566 | 1.183 | 10.276 |
| ATOM | 705 | C | LEU | A | 76 | -8.848 | -0.211 | 10.808 |
| ATOM | 706 | O | LEU | A | 76 | -8.514 | -0.582 | 11.958 |
| ATOM | 707 | CB | LEU | A | 76 | -7.103 | 1.270 | 9.798 |
| ATOM | 708 | CG | LEU | A | 76 | -6.608 | 2.684 | 9.443 |
| ATOM | 709 | CD1 | LEU | A | 76 | -5.151 | 2.645 | 9.087 |
| ATOM | 710 | CD2 | LEU | A | 76 | -7.396 | 3.302 | 8.296 |
| ATOM | 711 | N | VAL | A | 77 | -9.569 | -1.062 | 10.042 |
| ATOM | 712 | H | VAL | A | 77 | -9.894 | -0.766 | 9.144 |
| ATOM | 713 | CA | VAL | A | 77 | -9.899 | -2.428 | 10.485 |
| ATOM | 714 | C | VAL | A | 77 | -9.298 | -3.412 | 9.482 |
| ATOM | 715 | O | VAL | A | 77 | -9.450 | -3.300 | 8.253 |
| ATOM | 716 | CB | VAL | A | 77 | -11.436 | -2.592 | 10.506 |
| ATOM | 717 | CG1 | VAL | A | 77 | -11.830 | -4.021 | 10.682 |
| ATOM | 718 | CG2 | VAL | A | 77 | -12.072 | -1.765 | 11.634 |
| ATOM | 719 | N | GLY | A | 78 | -8.560 | -4.402 | 9.928 |
| ATOM | 720 | H | GLY | A | 78 | -8.445 | -4.530 | 10.913 |
| ATOM | 721 | CA | GLY | A | 78 | -7.930 | -5.285 | 8.987 |
| ATOM | 722 | C | GLY | A | 78 | -7.228 | -6.380 | 9.732 |
| ATOM | 723 | O | GLY | A | 78 | -7.292 | -6.524 | 10.970 |
| ATOM | 724 | N | PRO | A | 79 | -6.512 | -7.271 | 9.003 |
| ATOM | 725 | CA | PRO | A | 79 | -5.880 | -8.467 | 9.602 |
| ATOM | 726 | C | PRO | A | 79 | -4.599 | -8.107 | 10.340 |
| ATOM | 727 | O | PRO | A | 79 | -3.449 | -8.489 | 10.032 |

FIG. 11M

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 728 | CB | PRO | A | 79 | -5.613 | -9.379 | 8.400 |
| ATOM | 729 | CG | PRO | A | 79 | -5.529 | -8.416 | 7.210 |
| ATOM | 730 | CD | PRO | A | 79 | -6.415 | -7.225 | 7.537 |
| ATOM | 731 | N | THR | A | 80 | -4.759 | -7.304 | 11.408 |
| ATOM | 732 | H | THR | A | 80 | -5.664 | -6.935 | 11.619 |
| ATOM | 733 | CA | THR | A | 80 | -3.658 | -6.957 | 12.263 |
| ATOM | 734 | C | THR | A | 80 | -3.490 | -8.075 | 13.308 |
| ATOM | 735 | O | THR | A | 80 | -4.447 | -8.642 | 13.857 |
| ATOM | 736 | CB | THR | A | 80 | -3.868 | -5.572 | 12.927 |
| ATOM | 737 | OG1 | THR | A | 80 | -2.770 | -5.303 | 13.787 |
| ATOM | 738 | HG1 | THR | A | 80 | -2.889 | -4.412 | 14.225 |
| ATOM | 739 | CG2 | THR | A | 80 | -5.210 | -5.464 | 13.678 |
| ATOM | 740 | N | PRO | A | 81 | -2.243 | -8.496 | 13.589 |
| ATOM | 741 | CA | PRO | A | 81 | -1.986 | -9.476 | 14.660 |
| ATOM | 742 | C | PRO | A | 81 | -2.499 | -8.952 | 16.001 |
| ATOM | 743 | O | PRO | A | 81 | -2.944 | -9.720 | 16.866 |
| ATOM | 744 | CB | PRO | A | 81 | -0.444 | -9.549 | 14.732 |
| ATOM | 745 | CG | PRO | A | 81 | 0.069 | -8.951 | 13.429 |
| ATOM | 746 | CD | PRO | A | 81 | -1.029 | -8.105 | 12.842 |
| ATOM | 747 | N | VAL | A | 82 | -2.474 | -7.621 | 16.276 |
| ATOM | 748 | H | VAL | A | 82 | -2.180 | -6.975 | 15.571 |
| ATOM | 749 | CA | VAL | A | 82 | -2.869 | -7.091 | 17.591 |
| ATOM | 750 | C | VAL | A | 82 | -3.605 | -5.761 | 17.379 |
| ATOM | 751 | O | VAL | A | 82 | -3.349 | -5.004 | 16.429 |
| ATOM | 752 | CB | VAL | A | 82 | -1.595 | -6.858 | 18.443 |
| ATOM | 753 | CG1 | VAL | A | 82 | -0.650 | -5.824 | 17.803 |
| ATOM | 754 | CG2 | VAL | A | 82 | -1.907 | -6.418 | 19.890 |
| ATOM | 755 | N | ASN | A | 83 | -4.548 | -5.371 | 18.260 |
| ATOM | 756 | H | ASN | A | 83 | -4.810 | -5.981 | 19.007 |
| ATOM | 757 | CA | ASN | A | 83 | -5.181 | -4.067 | 18.123 |
| ATOM | 758 | C | ASN | A | 83 | -4.195 | -3.019 | 18.565 |
| ATOM | 759 | O | ASN | A | 83 | -3.605 | -3.064 | 19.665 |
| ATOM | 760 | CB | ASN | A | 83 | -6.436 | -3.942 | 18.982 |
| ATOM | 761 | CG | ASN | A | 83 | -7.502 | -4.930 | 18.631 |
| ATOM | 762 | OD1 | ASN | A | 83 | -7.899 | -5.049 | 17.488 |
| ATOM | 763 | ND2 | ASN | A | 83 | -7.980 | -5.662 | 19.628 |
| ATOM | 764 | 2HD2 | ASN | A | 83 | -8.695 | -6.341 | 19.459 |
| ATOM | 765 | 1HD2 | ASN | A | 83 | -7.630 | -5.541 | 20.557 |
| ATOM | 766 | N | ILE | A | 84 | -4.007 | -1.951 | 17.770 |
| ATOM | 767 | H | ILE | A | 84 | -4.583 | -1.827 | 16.962 |
| ATOM | 768 | CA | ILE | A | 84 | -2.993 | -0.954 | 18.032 |
| ATOM | 769 | C | ILE | A | 84 | -3.679 | 0.387 | 18.114 |
| ATOM | 770 | O | ILE | A | 84 | -4.460 | 0.797 | 17.240 |
| ATOM | 771 | CB | ILE | A | 84 | -2.021 | -0.922 | 16.833 |
| ATOM | 772 | CG1 | ILE | A | 84 | -1.162 | -2.150 | 16.859 |
| ATOM | 773 | CG2 | ILE | A | 84 | -1.219 | 0.387 | 16.747 |
| ATOM | 774 | CD1 | ILE | A | 84 | -0.375 | -2.360 | 15.579 |
| ATOM | 775 | N | ILE | A | 85 | -3.471 | 1.155 | 19.203 |
| ATOM | 776 | H | ILE | A | 85 | -2.972 | 0.781 | 19.985 |
| ATOM | 777 | CA | ILE | A | 85 | -3.951 | 2.518 | 19.281 |
| ATOM | 778 | C | ILE | A | 85 | -2.784 | 3.425 | 18.949 |
| ATOM | 779 | O | ILE | A | 85 | -1.767 | 3.515 | 19.663 |
| ATOM | 780 | CB | ILE | A | 85 | -4.522 | 2.825 | 20.676 |
| ATOM | 781 | CG1 | ILE | A | 85 | -5.673 | 1.865 | 21.050 |
| ATOM | 782 | CG2 | ILE | A | 85 | -5.000 | 4.274 | 20.716 |
| ATOM | 783 | CD1 | ILE | A | 85 | -6.828 | 1.808 | 20.059 |

FIG. 1 IN

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 784 | N | GLY | A | 86 | -2.820 | 4.123 | 17.792 |
| ATOM | 785 | H | GLY | A | 86 | -3.637 | 4.087 | 17.217 |
| ATOM | 786 | CA | GLY | A | 86 | -1.690 | 4.936 | 17.351 |
| ATOM | 787 | C | GLY | A | 86 | -1.831 | 6.393 | 17.704 |
| ATOM | 788 | O | GLY | A | 86 | -2.760 | 6.864 | 18.390 |
| ATOM | 789 | N | ARG | A | 87 | -0.881 | 7.229 | 17.230 |
| ATOM | 790 | H | ARG | A | 87 | -0.204 | 6.890 | 16.577 |
| ATOM | 791 | CA | ARG | A | 87 | -0.810 | 8.623 | 17.643 |
| ATOM | 792 | C | ARG | A | 87 | -2.027 | 9.445 | 17.277 |
| ATOM | 793 | O | ARG | A | 87 | -2.365 | 10.430 | 17.963 |
| ATOM | 794 | CB | ARG | A | 87 | 0.450 | 9.275 | 17.057 |
| ATOM | 795 | CG | ARG | A | 87 | 1.735 | 8.496 | 17.205 |
| ATOM | 796 | CD | ARG | A | 87 | 2.762 | 8.916 | 16.207 |
| ATOM | 797 | NE | ARG | A | 87 | 3.875 | 7.961 | 16.117 |
| ATOM | 798 | HE | ARG | A | 87 | 4.035 | 7.353 | 16.895 |
| ATOM | 799 | CZ | ARG | A | 87 | 4.660 | 7.893 | 15.035 |
| ATOM | 800 | NH1 | ARG | A | 87 | 4.463 | 8.675 | 13.975 |
| ATOM | 801 | 2HH1 | ARG | A | 87 | 3.712 | 9.335 | 13.974 |
| ATOM | 802 | 1HH1 | ARG | A | 87 | 5.066 | 8.602 | 13.181 |
| ATOM | 803 | NH2 | ARG | A | 87 | 5.656 | 7.019 | 15.023 |
| ATOM | 804 | 1HH2 | ARG | A | 87 | 6.254 | 6.953 | 14.224 |
| ATOM | 805 | 2HH2 | ARG | A | 87 | 5.810 | 6.426 | 15.813 |
| ATOM | 806 | N | ASN | A | 88 | -2.780 | 9.120 | 16.214 |
| ATOM | 807 | H | ASN | A | 88 | -2.504 | 8.361 | 15.625 |
| ATOM | 808 | CA | ASN | A | 88 | -4.015 | 9.860 | 15.890 |
| ATOM | 809 | C | ASN | A | 88 | -4.963 | 9.921 | 17.069 |
| ATOM | 810 | O | ASN | A | 88 | -5.613 | 10.954 | 17.345 |
| ATOM | 811 | CB | ASN | A | 88 | -4.712 | 9.315 | 14.617 |
| ATOM | 812 | CG | ASN | A | 88 | -5.475 | 8.001 | 14.827 |
| ATOM | 813 | OD1 | ASN | A | 88 | -4.922 | 6.996 | 15.245 |
| ATOM | 814 | ND2 | ASN | A | 88 | -6.758 | 7.998 | 14.506 |
| ATOM | 815 | 2HD2 | ASN | A | 88 | -7.306 | 7.169 | 14.622 |
| ATOM | 816 | 1HD2 | ASN | A | 88 | -7.190 | 8.824 | 14.145 |
| ATOM | 817 | N | LEU | A | 89 | -5.130 | 8.847 | 17.848 |
| ATOM | 818 | H | LEU | A | 89 | -4.637 | 8.002 | 17.640 |
| ATOM | 819 | CA | LEU | A | 89 | -6.024 | 8.865 | 19.013 |
| ATOM | 820 | C | LEU | A | 89 | -5.275 | 9.091 | 20.309 |
| ATOM | 821 | O | LEU | A | 89 | -5.834 | 9.632 | 21.283 |
| ATOM | 822 | CB | LEU | A | 89 | -6.840 | 7.592 | 19.140 |
| ATOM | 823 | CG | LEU | A | 89 | -7.759 | 7.355 | 17.957 |
| ATOM | 824 | CD1 | LEU | A | 89 | -8.369 | 5.980 | 18.088 |
| ATOM | 825 | CD2 | LEU | A | 89 | -8.817 | 8.457 | 17.801 |
| ATOM | 826 | N | LEU | A | 90 | -3.983 | 8.745 | 20.428 |
| ATOM | 827 | H | LEU | A | 90 | -3.525 | 8.274 | 19.674 |
| ATOM | 828 | CA | LEU | A | 90 | -3.242 | 9.057 | 21.664 |
| ATOM | 829 | C | LEU | A | 90 | -3.155 | 10.555 | 21.932 |
| ATOM | 830 | O | LEU | A | 90 | -3.202 | 11.020 | 23.092 |
| ATOM | 831 | CB | LEU | A | 90 | -1.817 | 8.453 | 21.661 |
| ATOM | 832 | CG | LEU | A | 90 | -1.766 | 6.914 | 21.587 |
| ATOM | 833 | CD1 | LEU | A | 90 | -0.343 | 6.494 | 21.396 |
| ATOM | 834 | CD2 | LEU | A | 90 | -2.339 | 6.230 | 22.812 |
| ATOM | 835 | N | THR | A | 91 | -3.031 | 11.407 | 20.926 |
| ATOM | 836 | H | THR | A | 91 | -2.982 | 11.063 | 19.988 |
| ATOM | 837 | CA | THR | A | 91 | -2.964 | 12.834 | 21.155 |
| ATOM | 838 | C | THR | A | 91 | -4.309 | 13.331 | 21.635 |
| ATOM | 839 | O | THR | A | 91 | -4.422 | 14.315 | 22.398 |

FIG. 1I0

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 840 | CB | THR | A | 91 | -2.555 | 13.543 | 19.848 |
| ATOM | 841 | OG1 | THR | A | 91 | -3.459 | 13.214 | 18.802 |
| ATOM | 842 | HG1 | THR | A | 91 | -3.188 | 13.677 | 17.958 |
| ATOM | 843 | CG2 | THR | A | 91 | -1.153 | 13.122 | 19.395 |
| ATOM | 844 | N | GLN | A | 92 | -5.435 | 12.704 | 21.258 |
| ATOM | 845 | H | GLN | A | 92 | -5.379 | 11.892 | 20.677 |
| ATOM | 846 | CA | GLN | A | 92 | -6.763 | 13.186 | 21.682 |
| ATOM | 847 | C | GLN | A | 92 | -6.942 | 12.975 | 23.153 |
| ATOM | 848 | O | GLN | A | 92 | -7.554 | 13.797 | 23.871 |
| ATOM | 849 | CB | GLN | A | 92 | -7.890 | 12.479 | 20.964 |
| ATOM | 850 | CG | GLN | A | 92 | -7.937 | 12.862 | 19.517 |
| ATOM | 851 | CD | GLN | A | 92 | -9.251 | 12.515 | 18.886 |
| ATOM | 852 | OE1 | GLN | A | 92 | -10.270 | 12.424 | 19.546 |
| ATOM | 853 | NE2 | GLN | A | 92 | -9.202 | 12.323 | 17.588 |
| ATOM | 854 | 1HE2 | GLN | A | 92 | -10.031 | 12.087 | 17.080 |
| ATOM | 855 | 2HE2 | GLN | A | 92 | -8.336 | 12.411 | 17.097 |
| ATOM | 856 | N | ILE | A | 93 | -6.472 | 11.846 | 23.721 |
| ATOM | 857 | H | ILE | A | 93 | -6.014 | 11.160 | 23.155 |
| ATOM | 858 | CA | ILE | A | 93 | -6.608 | 11.578 | 25.165 |
| ATOM | 859 | C | ILE | A | 93 | -5.472 | 12.189 | 25.948 |
| ATOM | 860 | O | ILE | A | 93 | -5.342 | 12.031 | 27.171 |
| ATOM | 861 | CB | ILE | A | 93 | -6.820 | 10.073 | 25.484 |
| ATOM | 862 | CG1 | ILE | A | 93 | -5.536 | 9.221 | 25.286 |
| ATOM | 863 | CG2 | ILE | A | 93 | -8.022 | 9.486 | 24.735 |
| ATOM | 864 | CD1 | ILE | A | 93 | -5.754 | 7.740 | 25.693 |
| ATOM | 865 | N | GLY | A | 94 | -4.594 | 12.993 | 25.330 |
| ATOM | 866 | H | GLY | A | 94 | -4.617 | 13.079 | 24.334 |
| ATOM | 867 | CA | GLY | A | 94 | -3.613 | 13.742 | 26.063 |
| ATOM | 868 | C | GLY | A | 94 | -2.448 | 12.895 | 26.512 |
| ATOM | 869 | O | GLY | A | 94 | -1.764 | 13.158 | 27.519 |
| ATOM | 870 | N | CYS | A | 95 | -2.117 | 11.849 | 25.797 |
| ATOM | 871 | H | CYS | A | 95 | -2.619 | 11.644 | 24.957 |
| ATOM | 872 | CA | CYS | A | 95 | -1.036 | 10.994 | 26.214 |
| ATOM | 873 | C | CYS | A | 95 | 0.362 | 11.566 | 25.925 |
| ATOM | 874 | O | CYS | A | 95 | 0.588 | 12.254 | 24.907 |
| ATOM | 875 | CB | CYS | A | 95 | -1.260 | 9.655 | 25.550 |
| ATOM | 876 | SG | CYS | A | 95 | -0.254 | 8.307 | 26.125 |
| ATOM | 877 | N | THR | A | 96 | 1.346 | 11.297 | 26.803 |
| ATOM | 878 | H | THR | A | 96 | 1.135 | 10.738 | 27.618 |
| ATOM | 879 | CA | THR | A | 96 | 2.728 | 11.779 | 26.664 |
| ATOM | 880 | C | THR | A | 96 | 3.729 | 10.784 | 27.264 |
| ATOM | 881 | O | THR | A | 96 | 3.498 | 10.249 | 28.345 |
| ATOM | 882 | CB | THR | A | 96 | 2.925 | 13.154 | 27.346 |
| ATOM | 883 | OG1 | THR | A | 96 | 2.594 | 13.109 | 28.721 |
| ATOM | 884 | HG1 | THR | A | 96 | 2.784 | 13.966 | 29.109 |
| ATOM | 885 | CG2 | THR | A | 96 | 2.139 | 14.300 | 26.698 |
| ATOM | 886 | N | LEU | A | 97 | 4.882 | 10.603 | 26.599 |
| ATOM | 887 | H | LEU | A | 97 | 5.016 | 11.071 | 25.714 |
| ATOM | 888 | CA | LEU | A | 97 | 6.040 | 9.910 | 27.166 |
| ATOM | 889 | C | LEU | A | 97 | 6.751 | 10.824 | 28.175 |
| ATOM | 890 | O | LEU | A | 97 | 6.705 | 12.046 | 28.044 |
| ATOM | 891 | CB | LEU | A | 97 | 7.013 | 9.497 | 26.049 |
| ATOM | 892 | CG | LEU | A | 97 | 6.452 | 8.449 | 25.065 |
| ATOM | 893 | CD1 | LEU | A | 97 | 7.360 | 8.355 | 23.828 |
| ATOM | 894 | CD2 | LEU | A | 97 | 6.345 | 7.065 | 25.724 |
| ATOM | 895 | N | ASN | A | 98 | 7.412 | 10.221 | 29.175 |

FIG. 11P

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 896 | H | ASN | A | 98 | 7.413 | 9.212 | 29.205 |
| ATOM | 897 | CA | ASN | A | 98 | 8.065 | 10.897 | 30.292 |
| ATOM | 898 | C | ASN | A | 98 | 9.220 | 10.029 | 30.800 |
| ATOM | 899 | O | ASN | A | 98 | 8.995 | 9.079 | 31.550 |
| ATOM | 900 | CB | ASN | A | 98 | 7.057 | 11.177 | 31.423 |
| ATOM | 901 | CG | ASN | A | 98 | 6.084 | 12.305 | 31.083 |
| ATOM | 902 | OD1 | ASN | A | 98 | 4.983 | 12.062 | 30.594 |
| ATOM | 903 | ND2 | ASN | A | 98 | 6.493 | 13.549 | 31.342 |
| ATOM | 904 | 2HD2 | ASN | A | 98 | 5.888 | 14.331 | 31.136 |
| ATOM | 905 | 1HD2 | ASN | A | 98 | 7.406 | 13.707 | 31.742 |
| ATOM | 906 | N | LEU | A | 99 | 10.451 | 10.369 | 30.389 |
| ATOM | 907 | H | LEU | A | 99 | 10.547 | 11.177 | 29.792 |
| ATOM | 908 | CA | LEU | A | 99 | 11.679 | 9.620 | 30.666 |
| ATOM | 909 | C | LEU | A | 99 | 12.711 | 10.437 | 31.454 |
| ATOM | 910 | O | LEU | A | 99 | 12.487 | 11.652 | 31.651 |
| ATOM | 911 | CB | LEU | A | 99 | 12.233 | 8.989 | 29.369 |
| ATOM | 912 | CG | LEU | A | 99 | 12.833 | 9.873 | 28.248 |
| ATOM | 913 | CD1 | LEU | A | 99 | 11.876 | 10.947 | 27.705 |
| ATOM | 914 | CD2 | LEU | A | 99 | 14.183 | 10.505 | 28.623 |
| ATOM | 915 | OXT | LEU | A | 99 | 13.716 | 9.819 | 31.869 |
| TER | | | | | | | | |
| ATOM | 916 | N | PRO | B | 1 | 12.600 | 14.237 | 30.106 |
| ATOM | 917 | CA | PRO | B | 1 | 11.842 | 15.268 | 29.363 |
| ATOM | 918 | C | PRO | B | 1 | 10.430 | 14.773 | 29.138 |
| ATOM | 919 | O | PRO | B | 1 | 10.054 | 13.695 | 29.618 |
| ATOM | 920 | CB | PRO | B | 1 | 12.622 | 15.412 | 28.035 |
| ATOM | 921 | CG | PRO | B | 1 | 13.817 | 14.470 | 28.131 |
| ATOM | 922 | CD | PRO | B | 1 | 13.966 | 14.227 | 29.603 |
| ATOM | 923 | 1H | PRO | B | 1 | 12.175 | 13.343 | 29.964 |
| ATOM | 924 | 2H | PRO | B | 1 | 12.594 | 14.457 | 31.081 |
| ATOM | 925 | N | GLN | B | 2 | 9.513 | 15.542 | 28.523 |
| ATOM | 926 | H | GLN | B | 2 | 9.751 | 16.474 | 28.251 |
| ATOM | 927 | CA | GLN | B | 2 | 8.186 | 15.058 | 28.242 |
| ATOM | 928 | C | GLN | B | 2 | 8.066 | 15.151 | 26.749 |
| ATOM | 929 | O | GLN | B | 2 | 8.523 | 16.140 | 26.133 |
| ATOM | 930 | CB | GLN | B | 2 | 7.155 | 15.976 | 28.856 |
| ATOM | 931 | CG | GLN | B | 2 | 5.739 | 15.732 | 28.373 |
| ATOM | 932 | CD | GLN | B | 2 | 4.744 | 16.365 | 29.284 |
| ATOM | 933 | OE1 | GLN | B | 2 | 4.628 | 15.962 | 30.431 |
| ATOM | 934 | NE2 | GLN | B | 2 | 4.024 | 17.367 | 28.784 |
| ATOM | 935 | 1HE2 | GLN | B | 2 | 3.341 | 17.830 | 29.349 |
| ATOM | 936 | 2HE2 | GLN | B | 2 | 4.160 | 17.665 | 27.839 |
| ATOM | 937 | N | ILE | B | 3 | 7.499 | 14.176 | 26.036 |
| ATOM | 938 | H | ILE | B | 3 | 7.102 | 13.386 | 26.504 |
| ATOM | 939 | CA | ILE | B | 3 | 7.435 | 14.216 | 24.601 |
| ATOM | 940 | C | ILE | B | 3 | 5.956 | 14.097 | 24.184 |
| ATOM | 941 | O | ILE | B | 3 | 5.150 | 13.290 | 24.710 |
| ATOM | 942 | CB | ILE | B | 3 | 8.299 | 13.058 | 24.029 |
| ATOM | 943 | CG1 | ILE | B | 3 | 9.743 | 13.232 | 24.534 |
| ATOM | 944 | CG2 | ILE | B | 3 | 8.269 | 12.985 | 22.496 |
| ATOM | 945 | CD1 | ILE | B | 3 | 10.621 | 12.068 | 24.143 |
| ATOM | 946 | N | THR | B | 4 | 5.462 | 15.108 | 23.453 |
| ATOM | 947 | H | THR | B | 4 | 6.046 | 15.887 | 23.226 |
| ATOM | 948 | CA | THR | B | 4 | 4.107 | 15.115 | 22.976 |
| ATOM | 949 | C | THR | B | 4 | 4.039 | 14.193 | 21.765 |
| ATOM | 950 | O | THR | B | 4 | 5.066 | 13.755 | 21.203 |

FIG. 1 IQ

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 951 | CB | THR | B | 4 | 3.616 | 16.548 | 22.647 |
| ATOM | 952 | OG1 | THR | B | 4 | 4.450 | 17.157 | 21.645 |
| ATOM | 953 | HG1 | THR | B | 4 | 4.123 | 18.080 | 21.442 |
| ATOM | 954 | CG2 | THR | B | 4 | 3.644 | 17.454 | 23.876 |
| ATOM | 955 | N | LEU | B | 5 | 2.872 | 13.781 | 21.324 |
| ATOM | 956 | H | LEU | B | 5 | 2.033 | 14.151 | 21.723 |
| ATOM | 957 | CA | LEU | B | 5 | 2.837 | 12.795 | 20.265 |
| ATOM | 958 | C | LEU | B | 5 | 2.183 | 13.415 | 19.047 |
| ATOM | 959 | O | LEU | B | 5 | 1.677 | 12.720 | 18.142 |
| ATOM | 960 | CB | LEU | B | 5 | 2.093 | 11.577 | 20.762 |
| ATOM | 961 | CG | LEU | B | 5 | 2.819 | 10.856 | 21.892 |
| ATOM | 962 | CD1 | LEU | B | 5 | 1.889 | 9.885 | 22.602 |
| ATOM | 963 | CD2 | LEU | B | 5 | 4.108 | 10.159 | 21.416 |
| ATOM | 964 | N | TRP | B | 6 | 2.209 | 14.742 | 18.880 |
| ATOM | 965 | H | TRP | B | 6 | 2.601 | 15.323 | 19.593 |
| ATOM | 966 | CA | TRP | B | 6 | 1.683 | 15.364 | 17.690 |
| ATOM | 967 | C | TRP | B | 6 | 2.581 | 14.978 | 16.509 |
| ATOM | 968 | O | TRP | B | 6 | 2.159 | 14.851 | 15.349 |
| ATOM | 969 | CB | TRP | B | 6 | 1.587 | 16.879 | 17.833 |
| ATOM | 970 | CG | TRP | B | 6 | 0.652 | 17.339 | 18.921 |
| ATOM | 971 | CD1 | TRP | B | 6 | 0.955 | 17.584 | 20.232 |
| ATOM | 972 | CD2 | TRP | B | 6 | -0.750 | 17.612 | 18.783 |
| ATOM | 973 | NE1 | TRP | B | 6 | -0.167 | 17.989 | 20.913 |
| ATOM | 974 | HE1 | TRP | B | 6 | -0.217 | 18.230 | 21.882 |
| ATOM | 975 | CE2 | TRP | B | 6 | -1.224 | 18.013 | 20.048 |
| ATOM | 976 | CE3 | TRP | B | 6 | -1.637 | 17.550 | 17.709 |
| ATOM | 977 | CZ2 | TRP | B | 6 | -2.544 | 18.352 | 20.266 |
| ATOM | 978 | CZ3 | TRP | B | 6 | -2.947 | 17.885 | 17.921 |
| ATOM | 979 | CH2 | TRP | B | 6 | -3.394 | 18.281 | 19.185 |
| ATOM | 980 | N | GLN | B | 7 | 3.896 | 14.809 | 16.738 |
| ATOM | 981 | H | GLN | B | 7 | 4.267 | 14.985 | 17.650 |
| ATOM | 982 | CA | GLN | B | 7 | 4.794 | 14.376 | 15.689 |
| ATOM | 983 | C | GLN | B | 7 | 5.361 | 13.043 | 16.096 |
| ATOM | 984 | O | GLN | B | 7 | 5.221 | 12.586 | 17.243 |
| ATOM | 985 | CB | GLN | B | 7 | 5.880 | 15.430 | 15.505 |
| ATOM | 986 | CG | GLN | B | 7 | 5.353 | 16.704 | 14.804 |
| ATOM | 987 | CD | GLN | B | 7 | 6.197 | 17.912 | 15.137 |
| ATOM | 988 | OE1 | GLN | B | 7 | 7.400 | 17.802 | 15.404 |
| ATOM | 989 | NE2 | GLN | B | 7 | 5.553 | 19.083 | 15.121 |
| ATOM | 990 | 1HE2 | GLN | B | 7 | 6.040 | 19.931 | 15.330 |
| ATOM | 991 | 2HE2 | GLN | B | 7 | 4.579 | 19.121 | 14.900 |
| ATOM | 992 | N | ARG | B | 8 | 5.979 | 12.274 | 15.189 |
| ATOM | 993 | H | ARG | B | 8 | 6.073 | 12.597 | 14.247 |
| ATOM | 994 | CA | ARG | B | 8 | 6.505 | 10.985 | 15.573 |
| ATOM | 995 | C | ARG | B | 8 | 7.577 | 11.198 | 16.610 |
| ATOM | 996 | O | ARG | B | 8 | 8.395 | 12.130 | 16.515 |
| ATOM | 997 | CB | ARG | B | 8 | 7.092 | 10.238 | 14.384 |
| ATOM | 998 | CG | ARG | B | 8 | 6.132 | 10.018 | 13.237 |
| ATOM | 999 | CD | ARG | B | 8 | 6.802 | 9.402 | 12.046 |
| ATOM | 1000 | NE | ARG | B | 8 | 5.846 | 9.005 | 11.023 |
| ATOM | 1001 | HE | ARG | B | 8 | 4.872 | 9.080 | 11.237 |
| ATOM | 1002 | CZ | ARG | B | 8 | 6.217 | 8.552 | 9.828 |
| ATOM | 1003 | NH1 | ARG | B | 8 | 7.496 | 8.442 | 9.486 |
| ATOM | 1004 | 2HH1 | ARG | B | 8 | 8.211 | 8.703 | 10.134 |
| ATOM | 1005 | 1HH1 | ARG | B | 8 | 7.744 | 8.098 | 8.580 |
| ATOM | 1006 | NH2 | ARG | B | 8 | 5.279 | 8.202 | 8.952 |

FIG. 1IR

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1007 | 1HH2 | ARG | B | 8 | 5.540 | 7.860 | 8.050 |
| ATOM | 1008 | 2HH2 | ARG | B | 8 | 4.312 | 8.281 | 9.196 |
| ATOM | 1009 | N | PRO | B | 9 | 7.663 | 10.381 | 17.682 |
| ATOM | 1010 | CA | PRO | B | 9 | 8.666 | 10.587 | 18.746 |
| ATOM | 1011 | C | PRO | B | 9 | 10.065 | 10.196 | 18.315 |
| ATOM | 1012 | O | PRO | B | 9 | 10.678 | 9.215 | 18.778 |
| ATOM | 1013 | CB | PRO | B | 9 | 8.148 | 9.682 | 19.878 |
| ATOM | 1014 | CG | PRO | B | 9 | 7.315 | 8.607 | 19.206 |
| ATOM | 1015 | CD | PRO | B | 9 | 6.708 | 9.323 | 18.004 |
| ATOM | 1016 | N | LEU | B | 10 | 10.685 | 10.969 | 17.400 |
| ATOM | 1017 | H | LEU | B | 10 | 10.201 | 11.746 | 16.998 |
| ATOM | 1018 | CA | LEU | B | 10 | 12.040 | 10.706 | 16.978 |
| ATOM | 1019 | C | LEU | B | 10 | 12.976 | 11.498 | 17.850 |
| ATOM | 1020 | O | LEU | B | 10 | 12.880 | 12.733 | 18.018 |
| ATOM | 1021 | CB | LEU | B | 10 | 12.250 | 11.170 | 15.554 |
| ATOM | 1022 | CG | LEU | B | 10 | 11.427 | 10.386 | 14.551 |
| ATOM | 1023 | CD1 | LEU | B | 10 | 11.385 | 11.175 | 13.276 |
| ATOM | 1024 | CD2 | LEU | B | 10 | 11.956 | 8.947 | 14.355 |
| ATOM | 1025 | N | VAL | B | 11 | 14.030 | 10.843 | 18.384 |
| ATOM | 1026 | H | VAL | B | 11 | 14.148 | 9.866 | 18.206 |
| ATOM | 1027 | CA | VAL | B | 11 | 15.018 | 11.517 | 19.223 |
| ATOM | 1028 | C | VAL | B | 11 | 16.400 | 11.111 | 18.740 |
| ATOM | 1029 | O | VAL | B | 11 | 16.581 | 10.201 | 17.911 |
| ATOM | 1030 | CB | VAL | B | 11 | 14.857 | 11.100 | 20.699 |
| ATOM | 1031 | CG1 | VAL | B | 11 | 13.514 | 11.586 | 21.293 |
| ATOM | 1032 | CG2 | VAL | B | 11 | 15.038 | 9.573 | 20.903 |
| ATOM | 1033 | N | THR | B | 12 | 17.485 | 11.739 | 19.232 |
| ATOM | 1034 | H | THR | B | 12 | 17.370 | 12.507 | 19.862 |
| ATOM | 1035 | CA | THR | B | 12 | 18.843 | 11.325 | 18.868 |
| ATOM | 1036 | C | THR | B | 12 | 19.377 | 10.284 | 19.837 |
| ATOM | 1037 | O | THR | B | 12 | 19.237 | 10.352 | 21.082 |
| ATOM | 1038 | CB | THR | B | 12 | 19.830 | 12.520 | 18.820 |
| ATOM | 1039 | OG1 | THR | B | 12 | 19.389 | 13.483 | 17.876 |
| ATOM | 1040 | HG1 | THR | B | 12 | 20.028 | 14.252 | 17.848 |
| ATOM | 1041 | CG2 | THR | B | 12 | 21.234 | 12.075 | 18.399 |
| ATOM | 1042 | N | ILE | B | 13 | 20.044 | 9.234 | 19.338 |
| ATOM | 1043 | H | ILE | B | 13 | 20.135 | 9.130 | 18.348 |
| ATOM | 1044 | CA | ILE | B | 13 | 20.641 | 8.239 | 20.176 |
| ATOM | 1045 | C | ILE | B | 13 | 22.119 | 8.226 | 19.855 |
| ATOM | 1046 | O | ILE | B | 13 | 22.579 | 8.817 | 18.865 |
| ATOM | 1047 | CB | ILE | B | 13 | 19.993 | 6.870 | 19.879 |
| ATOM | 1048 | CG1 | ILE | B | 13 | 20.192 | 6.464 | 18.415 |
| ATOM | 1049 | CG2 | ILE | B | 13 | 18.482 | 6.893 | 20.206 |
| ATOM | 1050 | CD1 | ILE | B | 13 | 19.829 | 5.035 | 18.106 |
| ATOM | 1051 | N | LYS | B | 14 | 22.973 | 7.618 | 20.661 |
| ATOM | 1052 | H | LYS | B | 14 | 22.652 | 7.243 | 21.531 |
| ATOM | 1053 | CA | LYS | B | 14 | 24.364 | 7.480 | 20.317 |
| ATOM | 1054 | C | LYS | B | 14 | 24.680 | 6.029 | 20.477 |
| ATOM | 1055 | O | LYS | B | 14 | 24.353 | 5.353 | 21.484 |
| ATOM | 1056 | CB | LYS | B | 14 | 25.266 | 8.263 | 21.242 |
| ATOM | 1057 | CG | LYS | B | 14 | 24.947 | 9.729 | 21.236 |
| ATOM | 1058 | CD | LYS | B | 14 | 25.664 | 10.498 | 22.339 |
| ATOM | 1059 | CE | LYS | B | 14 | 26.758 | 11.441 | 21.807 |
| ATOM | 1060 | NZ | LYS | B | 14 | 28.026 | 10.781 | 21.440 |
| ATOM | 1061 | 1HZ | LYS | B | 14 | 28.674 | 11.466 | 21.107 |
| ATOM | 1062 | 3HZ | LYS | B | 14 | 27.855 | 10.107 | 20.722 |

FIG. 11S

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1063 | 2HZ | LYS | B | 14 | 28.408 | 10.323 | 22.243 |
| ATOM | 1064 | N | ILE | B | 15 | 25.214 | 5.390 | 19.425 |
| ATOM | 1065 | H | ILE | B | 15 | 25.434 | 5.901 | 18.594 |
| ATOM | 1066 | CA | ILE | B | 15 | 25.489 | 3.989 | 19.434 |
| ATOM | 1067 | C | ILE | B | 15 | 26.832 | 3.981 | 18.750 |
| ATOM | 1068 | O | ILE | B | 15 | 27.104 | 4.869 | 17.933 |
| ATOM | 1069 | CB | ILE | B | 15 | 24.435 | 3.220 | 18.606 |
| ATOM | 1070 | CG1 | ILE | B | 15 | 24.893 | 1.824 | 18.347 |
| ATOM | 1071 | CG2 | ILE | B | 15 | 24.048 | 3.977 | 17.309 |
| ATOM | 1072 | CD1 | ILE | B | 15 | 23.830 | 0.996 | 17.645 |
| ATOM | 1073 | N | GLY | B | 16 | 27.812 | 3.212 | 19.202 |
| ATOM | 1074 | H | GLY | B | 16 | 27.623 | 2.535 | 19.913 |
| ATOM | 1075 | CA | GLY | B | 16 | 29.175 | 3.336 | 18.677 |
| ATOM | 1076 | C | GLY | B | 16 | 29.771 | 4.754 | 18.619 |
| ATOM | 1077 | O | GLY | B | 16 | 30.737 | 4.970 | 17.902 |
| ATOM | 1078 | N | GLY | B | 17 | 29.273 | 5.791 | 19.335 |
| ATOM | 1079 | H | GLY | B | 17 | 28.453 | 5.660 | 19.892 |
| ATOM | 1080 | CA | GLY | B | 17 | 29.924 | 7.105 | 19.302 |
| ATOM | 1081 | C | GLY | B | 17 | 29.468 | 8.043 | 18.176 |
| ATOM | 1082 | O | GLY | B | 17 | 29.984 | 9.155 | 17.933 |
| ATOM | 1083 | N | GLN | B | 18 | 28.433 | 7.621 | 17.411 |
| ATOM | 1084 | H | GLN | B | 18 | 28.046 | 6.711 | 17.560 |
| ATOM | 1085 | CA | GLN | B | 18 | 27.834 | 8.449 | 16.348 |
| ATOM | 1086 | C | GLN | B | 18 | 26.407 | 8.755 | 16.736 |
| ATOM | 1087 | O | GLN | B | 18 | 25.678 | 7.953 | 17.353 |
| ATOM | 1088 | CB | GLN | B | 18 | 27.810 | 7.645 | 15.045 |
| ATOM | 1089 | CG | GLN | B | 18 | 27.247 | 6.204 | 15.146 |
| ATOM | 1090 | CD | GLN | B | 18 | 27.572 | 5.333 | 13.924 |
| ATOM | 1091 | OE1 | GLN | B | 18 | 26.771 | 4.501 | 13.464 |
| ATOM | 1092 | NE2 | GLN | B | 18 | 28.766 | 5.531 | 13.393 |
| ATOM | 1093 | 1HE2 | GLN | B | 18 | 29.057 | 5.005 | 12.594 |
| ATOM | 1094 | 2HE2 | GLN | B | 18 | 29.388 | 6.209 | 13.786 |
| ATOM | 1095 | N | LEU | B | 19 | 25.873 | 9.933 | 16.337 |
| ATOM | 1096 | H | LEU | B | 19 | 26.446 | 10.602 | 15.863 |
| ATOM | 1097 | CA | LEU | B | 19 | 24.467 | 10.267 | 16.578 |
| ATOM | 1098 | C | LEU | B | 19 | 23.633 | 9.622 | 15.490 |
| ATOM | 1099 | O | LEU | B | 19 | 23.912 | 9.707 | 14.284 |
| ATOM | 1100 | CB | LEU | B | 19 | 24.207 | 11.777 | 16.457 |
| ATOM | 1101 | CG | LEU | B | 19 | 24.857 | 12.756 | 17.454 |
| ATOM | 1102 | CD1 | LEU | B | 19 | 24.739 | 12.335 | 18.880 |
| ATOM | 1103 | CD2 | LEU | B | 19 | 26.299 | 13.072 | 17.130 |
| ATOM | 1104 | N | LYS | B | 20 | 22.450 | 9.085 | 15.850 |
| ATOM | 1105 | H | LYS | B | 20 | 22.242 | 8.948 | 16.819 |
| ATOM | 1106 | CA | LYS | B | 20 | 21.472 | 8.702 | 14.867 |
| ATOM | 1107 | C | LYS | B | 20 | 20.121 | 9.105 | 15.417 |
| ATOM | 1108 | O | LYS | B | 20 | 19.957 | 9.572 | 16.569 |
| ATOM | 1109 | CB | LYS | B | 20 | 21.496 | 7.200 | 14.560 |
| ATOM | 1110 | CG | LYS | B | 20 | 22.904 | 6.653 | 14.507 |
| ATOM | 1111 | CD | LYS | B | 20 | 23.052 | 5.366 | 13.677 |
| ATOM | 1112 | CE | LYS | B | 20 | 23.069 | 5.603 | 12.145 |
| ATOM | 1113 | NZ | LYS | B | 20 | 23.893 | 6.758 | 11.699 |
| ATOM | 1114 | 1HZ | LYS | B | 20 | 23.847 | 6.836 | 10.703 |
| ATOM | 1115 | 3HZ | LYS | B | 20 | 24.843 | 6.617 | 11.978 |
| ATOM | 1116 | 2HZ | LYS | B | 20 | 23.544 | 7.597 | 12.116 |
| ATOM | 1117 | N | GLU | B | 21 | 19.068 | 9.022 | 14.591 |
| ATOM | 1118 | H | GLU | B | 21 | 19.200 | 8.712 | 13.650 |

FIG. 11T

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1119 | CA | GLU | B | 21 | 17.735 | 9.366 | 15.008 |
| ATOM | 1120 | C | GLU | B | 21 | 16.937 | 8.095 | 15.119 |
| ATOM | 1121 | O | GLU | B | 21 | 17.117 | 7.103 | 14.376 |
| ATOM | 1122 | CB | GLU | B | 21 | 17.143 | 10.314 | 13.983 |
| ATOM | 1123 | CG | GLU | B | 21 | 15.714 | 10.706 | 14.162 |
| ATOM | 1124 | CD | GLU | B | 21 | 15.304 | 11.607 | 13.036 |
| ATOM | 1125 | OE1 | GLU | B | 21 | 14.971 | 11.051 | 11.957 |
| ATOM | 1126 | OE2 | GLU | B | 21 | 15.338 | 12.854 | 13.174 |
| ATOM | 1127 | N | ALA | B | 22 | 16.025 | 7.999 | 16.072 |
| ATOM | 1128 | H | ALA | B | 22 | 15.825 | 8.792 | 16.648 |
| ATOM | 1129 | CA | ALA | B | 22 | 15.300 | 6.783 | 16.315 |
| ATOM | 1130 | C | ALA | B | 22 | 13.981 | 7.132 | 16.952 |
| ATOM | 1131 | O | ALA | B | 22 | 13.756 | 8.153 | 17.632 |
| ATOM | 1132 | CB | ALA | B | 22 | 16.095 | 5.865 | 17.235 |
| ATOM | 1133 | N | LEU | B | 23 | 12.994 | 6.230 | 16.743 |
| ATOM | 1134 | H | LEU | B | 23 | 13.195 | 5.379 | 16.257 |
| ATOM | 1135 | CA | LEU | B | 23 | 11.639 | 6.408 | 17.180 |
| ATOM | 1136 | C | LEU | B | 23 | 11.476 | 5.740 | 18.534 |
| ATOM | 1137 | O | LEU | B | 23 | 11.814 | 4.564 | 18.746 |
| ATOM | 1138 | CB | LEU | B | 23 | 10.775 | 5.665 | 16.192 |
| ATOM | 1139 | CG | LEU | B | 23 | 9.267 | 5.810 | 16.237 |
| ATOM | 1140 | CD1 | LEU | B | 23 | 8.807 | 7.142 | 15.664 |
| ATOM | 1141 | CD2 | LEU | B | 23 | 8.648 | 4.625 | 15.482 |
| ATOM | 1142 | N | LEU | B | 24 | 10.948 | 6.455 | 19.553 |
| ATOM | 1143 | H | LEU | B | 24 | 10.775 | 7.433 | 19.435 |
| ATOM | 1144 | CA | LEU | B | 24 | 10.613 | 5.838 | 20.849 |
| ATOM | 1145 | C | LEU | B | 24 | 9.271 | 5.160 | 20.687 |
| ATOM | 1146 | O | LEU | B | 24 | 8.208 | 5.764 | 20.418 |
| ATOM | 1147 | CB | LEU | B | 24 | 10.564 | 6.878 | 21.971 |
| ATOM | 1148 | CG | LEU | B | 24 | 11.828 | 7.750 | 22.075 |
| ATOM | 1149 | CD1 | LEU | B | 24 | 11.580 | 8.859 | 23.077 |
| ATOM | 1150 | CD2 | LEU | B | 24 | 13.099 | 6.955 | 22.388 |
| ATOM | 1151 | N | ASP | B | 25 | 9.246 | 3.822 | 20.809 |
| ATOM | 1152 | H | ASP | B | 25 | 10.025 | 3.347 | 21.218 |
| ATOM | 1153 | CA | ASP | B | 25 | 8.122 | 3.030 | 20.366 |
| ATOM | 1154 | C | ASP | B | 25 | 7.637 | 2.136 | 21.484 |
| ATOM | 1155 | O | ASP | B | 25 | 8.189 | 1.048 | 21.759 |
| ATOM | 1156 | CB | ASP | B | 25 | 8.613 | 2.196 | 19.189 |
| ATOM | 1157 | CG | ASP | B | 25 | 7.528 | 1.421 | 18.511 |
| ATOM | 1158 | OD1 | ASP | B | 25 | 6.422 | 1.339 | 19.058 |
| ATOM | 1159 | OD2 | ASP | B | 25 | 7.800 | 0.897 | 17.426 |
| ATOM | 1160 | N | THR | B | 26 | 6.547 | 2.465 | 22.157 |
| ATOM | 1161 | H | THR | B | 26 | 6.067 | 3.314 | 21.938 |
| ATOM | 1162 | CA | THR | B | 26 | 6.025 | 1.621 | 23.212 |
| ATOM | 1163 | C | THR | B | 26 | 5.347 | 0.369 | 22.694 |
| ATOM | 1164 | O | THR | B | 26 | 4.976 | -0.550 | 23.451 |
| ATOM | 1165 | CB | THR | B | 26 | 5.027 | 2.389 | 24.046 |
| ATOM | 1166 | OG1 | THR | B | 26 | 3.927 | 2.853 | 23.239 |
| ATOM | 1167 | HG1 | THR | B | 26 | 3.277 | 3.359 | 23.806 |
| ATOM | 1168 | CG2 | THR | B | 26 | 5.703 | 3.603 | 24.650 |
| ATOM | 1169 | N | GLY | B | 27 | 5.090 | 0.245 | 21.382 |
| ATOM | 1170 | H | GLY | B | 27 | 5.341 | 0.983 | 20.756 |
| ATOM | 1171 | CA | GLY | B | 27 | 4.457 | -0.938 | 20.867 |
| ATOM | 1172 | C | GLY | B | 27 | 5.475 | -1.992 | 20.458 |
| ATOM | 1173 | O | GLY | B | 27 | 5.121 | -3.108 | 20.055 |
| ATOM | 1174 | N | ALA | B | 28 | 6.792 | -1.717 | 20.495 |

FIG. 11U

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1175 | H | ALA | B | 28 | 7.104 | -0.832 | 20.841 |
| ATOM | 1176 | CA | ALA | B | 28 | 7.800 | -2.690 | 20.037 |
| ATOM | 1177 | C | ALA | B | 28 | 8.371 | -3.444 | 21.259 |
| ATOM | 1178 | O | ALA | B | 28 | 8.840 | -2.807 | 22.213 |
| ATOM | 1179 | CB | ALA | B | 28 | 8.924 | -1.936 | 19.358 |
| ATOM | 1180 | N | ASP | B | 29 | 8.459 | -4.787 | 21.289 |
| ATOM | 1181 | H | ASP | B | 29 | 8.082 | -5.325 | 20.535 |
| ATOM | 1182 | CA | ASP | B | 29 | 9.121 | -5.441 | 22.452 |
| ATOM | 1183 | C | ASP | B | 29 | 10.608 | -5.219 | 22.404 |
| ATOM | 1184 | O | ASP | B | 29 | 11.345 | -5.264 | 23.412 |
| ATOM | 1185 | CB | ASP | B | 29 | 8.965 | -6.975 | 22.447 |
| ATOM | 1186 | CG | ASP | B | 29 | 7.551 | -7.477 | 22.774 |
| ATOM | 1187 | OD1 | ASP | B | 29 | 6.683 | -6.693 | 23.169 |
| ATOM | 1188 | OD2 | ASP | B | 29 | 7.350 | -8.686 | 22.616 |
| ATOM | 1189 | N | ASP | B | 30 | 11.164 | -5.157 | 21.171 |
| ATOM | 1190 | H | ASP | B | 30 | 10.577 | -5.063 | 20.367 |
| ATOM | 1191 | CA | ASP | B | 30 | 12.609 | -5.217 | 20.880 |
| ATOM | 1192 | C | ASP | B | 30 | 13.048 | -3.886 | 20.335 |
| ATOM | 1193 | O | ASP | B | 30 | 12.269 | -3.055 | 19.817 |
| ATOM | 1194 | CB | ASP | B | 30 | 12.833 | -6.226 | 19.735 |
| ATOM | 1195 | CG | ASP | B | 30 | 12.477 | -7.675 | 20.099 |
| ATOM | 1196 | OD1 | ASP | B | 30 | 13.197 | -8.272 | 20.908 |
| ATOM | 1197 | OD2 | ASP | B | 30 | 11.494 | -8.237 | 19.569 |
| ATOM | 1198 | N | THR | B | 31 | 14.387 | -3.692 | 20.227 |
| ATOM | 1199 | H | THR | B | 31 | 15.018 | -4.380 | 20.586 |
| ATOM | 1200 | CA | THR | B | 31 | 14.981 | -2.530 | 19.614 |
| ATOM | 1201 | C | THR | B | 31 | 15.578 | -2.979 | 18.260 |
| ATOM | 1202 | O | THR | B | 31 | 16.246 | -4.020 | 18.123 |
| ATOM | 1203 | CB | THR | B | 31 | 16.036 | -2.004 | 20.557 |
| ATOM | 1204 | OG1 | THR | B | 31 | 15.378 | -1.376 | 21.645 |
| ATOM | 1205 | HG1 | THR | B | 31 | 16.052 | -1.016 | 22.290 |
| ATOM | 1206 | CG2 | THR | B | 31 | 16.944 | -0.960 | 19.904 |
| ATOM | 1207 | N | VAL | B | 32 | 15.237 | -2.283 | 17.150 |
| ATOM | 1208 | H | VAL | B | 32 | 14.703 | -1.442 | 17.237 |
| ATOM | 1209 | CA | VAL | B | 32 | 15.626 | -2.722 | 15.806 |
| ATOM | 1210 | C | VAL | B | 32 | 16.303 | -1.566 | 15.132 |
| ATOM | 1211 | O | VAL | B | 32 | 15.779 | -0.428 | 14.995 |
| ATOM | 1212 | CB | VAL | B | 32 | 14.407 | -3.126 | 14.964 |
| ATOM | 1213 | CG1 | VAL | B | 32 | 14.820 | -3.703 | 13.596 |
| ATOM | 1214 | CG2 | VAL | B | 32 | 13.556 | -4.102 | 15.703 |
| ATOM | 1215 | N | LEU | B | 33 | 17.563 | -1.756 | 14.720 |
| ATOM | 1216 | H | LEU | B | 33 | 17.984 | -2.658 | 14.814 |
| ATOM | 1217 | CA | LEU | B | 33 | 18.347 | -0.697 | 14.138 |
| ATOM | 1218 | C | LEU | B | 33 | 18.610 | -1.009 | 12.685 |
| ATOM | 1219 | O | LEU | B | 33 | 18.685 | -2.162 | 12.205 |
| ATOM | 1220 | CB | LEU | B | 33 | 19.679 | -0.628 | 14.856 |
| ATOM | 1221 | CG | LEU | B | 33 | 19.698 | 0.363 | 16.031 |
| ATOM | 1222 | CD1 | LEU | B | 33 | 18.425 | 0.321 | 16.891 |
| ATOM | 1223 | CD2 | LEU | B | 33 | 20.929 | 0.179 | 16.889 |
| ATOM | 1224 | N | GLU | B | 34 | 18.786 | 0.078 | 11.899 |
| ATOM | 1225 | H | GLU | B | 34 | 18.619 | 0.991 | 12.271 |
| ATOM | 1226 | CA | GLU | B | 34 | 19.218 | 0.041 | 10.488 |
| ATOM | 1227 | C | GLU | B | 34 | 20.478 | -0.774 | 10.399 |
| ATOM | 1228 | O | GLU | B | 34 | 21.374 | -0.835 | 11.272 |
| ATOM | 1229 | CB | GLU | B | 34 | 19.536 | 1.460 | 9.996 |
| ATOM | 1230 | CG | GLU | B | 34 | 20.722 | 2.088 | 10.761 |

FIG. 1IV

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1231 | CD | GLU | B | 34 | 21.085 | 3.512 | 10.314 |
| ATOM | 1232 | OE1 | GLU | B | 34 | 20.285 | 4.466 | 10.500 |
| ATOM | 1233 | OE2 | GLU | B | 34 | 22.211 | 3.703 | 9.775 |
| ATOM | 1234 | N | GLU | B | 35 | 20.673 | -1.367 | 9.205 |
| ATOM | 1235 | H | GLU | B | 35 | 20.011 | -1.227 | 8.468 |
| ATOM | 1236 | CA | GLU | B | 35 | 21.802 | -2.205 | 8.930 |
| ATOM | 1237 | C | GLU | B | 35 | 23.096 | -1.520 | 9.321 |
| ATOM | 1238 | O | GLU | B | 35 | 23.391 | -0.379 | 8.916 |
| ATOM | 1239 | CB | GLU | B | 35 | 21.741 | -2.479 | 7.439 |
| ATOM | 1240 | CG | GLU | B | 35 | 22.795 | -3.380 | 6.883 |
| ATOM | 1241 | CD | GLU | B | 35 | 22.987 | -4.587 | 7.744 |
| ATOM | 1242 | OE1 | GLU | B | 35 | 21.980 | -5.258 | 8.118 |
| ATOM | 1243 | OE2 | GLU | B | 35 | 24.149 | -4.860 | 8.048 |
| ATOM | 1244 | N | MET | B | 36 | 23.926 | -2.106 | 10.157 |
| ATOM | 1245 | H | MET | B | 36 | 23.654 | -2.953 | 10.613 |
| ATOM | 1246 | CA | MET | B | 36 | 25.232 | -1.559 | 10.441 |
| ATOM | 1247 | C | MET | B | 36 | 26.146 | -2.687 | 10.815 |
| ATOM | 1248 | O | MET | B | 36 | 25.731 | -3.783 | 11.257 |
| ATOM | 1249 | CB | MET | B | 36 | 25.251 | -0.424 | 11.497 |
| ATOM | 1250 | CG | MET | B | 36 | 24.626 | -0.724 | 12.881 |
| ATOM | 1251 | SD | MET | B | 36 | 24.722 | 0.719 | 13.988 |
| ATOM | 1252 | CE | MET | B | 36 | 23.132 | 1.586 | 13.692 |
| ATOM | 1253 | N | SER | B | 37 | 27.441 | -2.551 | 10.593 |
| ATOM | 1254 | H | SER | B | 37 | 27.783 | -1.726 | 10.144 |
| ATOM | 1255 | CA | SER | B | 37 | 28.321 | -3.608 | 11.011 |
| ATOM | 1256 | C | SER | B | 37 | 28.721 | -3.352 | 12.442 |
| ATOM | 1257 | O | SER | B | 37 | 29.402 | -2.369 | 12.788 |
| ATOM | 1258 | CB | SER | B | 37 | 29.567 | -3.622 | 10.109 |
| ATOM | 1259 | OG | SER | B | 37 | 29.231 | -3.908 | 8.750 |
| ATOM | 1260 | HG | SER | B | 37 | 30.057 | -3.911 | 8.187 |
| ATOM | 1261 | N | LEU | B | 38 | 28.469 | -4.295 | 13.366 |
| ATOM | 1262 | H | LEU | B | 38 | 27.948 | -5.123 | 13.117 |
| ATOM | 1263 | CA | LEU | B | 38 | 29.073 | -4.232 | 14.714 |
| ATOM | 1264 | C | LEU | B | 38 | 30.132 | -5.342 | 14.895 |
| ATOM | 1265 | O | LEU | B | 38 | 30.070 | -6.357 | 14.197 |
| ATOM | 1266 | CB | LEU | B | 38 | 27.986 | -4.237 | 15.802 |
| ATOM | 1267 | CG | LEU | B | 38 | 27.005 | -3.039 | 15.750 |
| ATOM | 1268 | CD1 | LEU | B | 38 | 25.885 | -3.214 | 16.788 |
| ATOM | 1269 | CD2 | LEU | B | 38 | 27.707 | -1.696 | 16.017 |
| ATOM | 1270 | N | PRO | B | 39 | 31.119 | -5.160 | 15.804 |
| ATOM | 1271 | CA | PRO | B | 39 | 32.199 | -6.116 | 16.052 |
| ATOM | 1272 | C | PRO | B | 39 | 31.767 | -7.223 | 17.028 |
| ATOM | 1273 | O | PRO | B | 39 | 31.448 | -6.942 | 18.185 |
| ATOM | 1274 | CB | PRO | B | 39 | 33.347 | -5.276 | 16.625 |
| ATOM | 1275 | CG | PRO | B | 39 | 32.634 | -4.148 | 17.370 |
| ATOM | 1276 | CD | PRO | B | 39 | 31.385 | -3.916 | 16.523 |
| ATOM | 1277 | N | GLY | B | 40 | 31.770 | -8.481 | 16.559 |
| ATOM | 1278 | H | GLY | B | 40 | 32.036 | -8.641 | 15.598 |
| ATOM | 1279 | CA | GLY | B | 40 | 31.420 | -9.658 | 17.353 |
| ATOM | 1280 | C | GLY | B | 40 | 30.679 | -10.723 | 16.539 |
| ATOM | 1281 | O | GLY | B | 40 | 30.647 | -10.671 | 15.308 |
| ATOM | 1282 | N | LYS | B | 41 | 30.098 | -11.699 | 17.255 |
| ATOM | 1283 | H | LYS | B | 41 | 30.164 | -11.656 | 18.261 |
| ATOM | 1284 | CA | LYS | B | 41 | 29.399 | -12.861 | 16.702 |
| ATOM | 1285 | C | LYS | B | 41 | 27.971 | -12.923 | 17.245 |
| ATOM | 1286 | O | LYS | B | 41 | 27.743 | -12.700 | 18.436 |

FIG. 11W

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1287 | CB | LYS | B | 41 | 30.154 | -14.152 | 17.048 |
| ATOM | 1288 | CG | LYS | B | 41 | 31.537 | -14.221 | 16.384 |
| ATOM | 1289 | CD | LYS | B | 41 | 32.192 | -15.580 | 16.651 |
| ATOM | 1290 | CE | LYS | B | 41 | 33.566 | -15.642 | 15.983 |
| ATOM | 1291 | NZ | LYS | B | 41 | 34.198 | -16.956 | 16.183 |
| ATOM | 1292 | 1HZ | LYS | B | 41 | 35.102 | -16.968 | 15.732 |
| ATOM | 1293 | 3HZ | LYS | B | 41 | 33.612 | -17.674 | 15.782 |
| ATOM | 1294 | 2HZ | LYS | B | 41 | 34.312 | -17.128 | 17.172 |
| ATOM | 1295 | N | TRP | B | 42 | 27.018 | -13.228 | 16.351 |
| ATOM | 1296 | H | TRP | B | 42 | 27.307 | -13.458 | 15.411 |
| ATOM | 1297 | CA | TRP | B | 42 | 25.597 | -12.929 | 16.521 |
| ATOM | 1298 | C | TRP | B | 42 | 24.723 | -14.179 | 16.405 |
| ATOM | 1299 | O | TRP | B | 42 | 25.210 | -15.277 | 16.131 |
| ATOM | 1300 | CB | TRP | B | 42 | 25.192 | -11.856 | 15.491 |
| ATOM | 1301 | CG | TRP | B | 42 | 26.127 | -10.687 | 15.390 |
| ATOM | 1302 | CD1 | TRP | B | 42 | 26.651 | -10.197 | 14.244 |
| ATOM | 1303 | CD2 | TRP | B | 42 | 26.739 | -9.913 | 16.467 |
| ATOM | 1304 | NE1 | TRP | B | 42 | 27.548 | -9.191 | 14.533 |
| ATOM | 1305 | HE1 | TRP | B | 42 | 28.067 | -8.702 | 13.818 |
| ATOM | 1306 | CE2 | TRP | B | 42 | 27.664 | -8.995 | 15.893 |
| ATOM | 1307 | CE3 | TRP | B | 42 | 26.640 | -9.923 | 17.875 |
| ATOM | 1308 | CZ2 | TRP | B | 42 | 28.443 | -8.136 | 16.680 |
| ATOM | 1309 | CZ3 | TRP | B | 42 | 27.426 | -9.075 | 18.673 |
| ATOM | 1310 | CH2 | TRP | B | 42 | 28.318 | -8.171 | 18.077 |
| ATOM | 1311 | N | LYS | B | 43 | 23.416 | -13.980 | 16.617 |
| ATOM | 1312 | H | LYS | B | 43 | 23.105 | -13.044 | 16.840 |
| ATOM | 1313 | CA | LYS | B | 43 | 22.378 | -14.995 | 16.526 |
| ATOM | 1314 | C | LYS | B | 43 | 21.368 | -14.507 | 15.478 |
| ATOM | 1315 | O | LYS | B | 43 | 20.743 | -13.472 | 15.706 |
| ATOM | 1316 | CB | LYS | B | 43 | 21.694 | -15.196 | 17.893 |
| ATOM | 1317 | CG | LYS | B | 43 | 22.641 | -15.623 | 19.034 |
| ATOM | 1318 | CD | LYS | B | 43 | 22.409 | -14.814 | 20.323 |
| ATOM | 1319 | CE | LYS | B | 43 | 22.767 | -13.327 | 20.182 |
| ATOM | 1320 | NZ | LYS | B | 43 | 24.214 | -13.113 | 20.015 |
| ATOM | 1321 | 1HZ | LYS | B | 43 | 24.400 | -12.125 | 19.924 |
| ATOM | 1322 | 3HZ | LYS | B | 43 | 24.532 | -13.593 | 19.185 |
| ATOM | 1323 | 2HZ | LYS | B | 43 | 24.702 | -13.476 | 20.821 |
| ATOM | 1324 | N | PRO | B | 44 | 21.175 | -15.204 | 14.341 |
| ATOM | 1325 | CA | PRO | B | 44 | 20.139 | -14.835 | 13.382 |
| ATOM | 1326 | C | PRO | B | 44 | 18.765 | -14.997 | 14.044 |
| ATOM | 1327 | O | PRO | B | 44 | 18.573 | -15.902 | 14.860 |
| ATOM | 1328 | CB | PRO | B | 44 | 20.341 | -15.761 | 12.180 |
| ATOM | 1329 | CG | PRO | B | 44 | 20.999 | -16.999 | 12.787 |
| ATOM | 1330 | CD | PRO | B | 44 | 21.837 | -16.434 | 13.933 |
| ATOM | 1331 | N | LYS | B | 45 | 17.825 | -14.101 | 13.712 |
| ATOM | 1332 | H | LYS | B | 45 | 17.994 | -13.483 | 12.944 |
| ATOM | 1333 | CA | LYS | B | 45 | 16.523 | -14.088 | 14.339 |
| ATOM | 1334 | C | LYS | B | 45 | 15.519 | -13.590 | 13.329 |
| ATOM | 1335 | O | LYS | B | 45 | 15.829 | -12.838 | 12.379 |
| ATOM | 1336 | CB | LYS | B | 45 | 16.558 | -13.149 | 15.560 |
| ATOM | 1337 | CG | LYS | B | 45 | 15.469 | -13.442 | 16.579 |
| ATOM | 1338 | CD | LYS | B | 45 | 15.256 | -12.254 | 17.501 |
| ATOM | 1339 | CE | LYS | B | 45 | 14.131 | -12.461 | 18.469 |
| ATOM | 1340 | NZ | LYS | B | 45 | 14.549 | -13.442 | 19.474 |
| ATOM | 1341 | 1HZ | LYS | B | 45 | 13.805 | -13.588 | 20.126 |
| ATOM | 1342 | 3HZ | LYS | B | 45 | 15.355 | -13.101 | 19.958 |

FIG. 11X

```
ATOM   1343  2HZ  LYS B  45      14.772 -14.306  19.023
ATOM   1344  N    MET B  46      14.240 -14.005  13.416
ATOM   1345  H    MET B  46      13.991 -14.705  14.085
ATOM   1346  CA   MET B  46      13.203 -13.472  12.570
ATOM   1347  C    MET B  46      12.291 -12.623  13.425
ATOM   1348  O    MET B  46      11.782 -13.063  14.471
ATOM   1349  CB   MET B  46      12.383 -14.616  12.016
ATOM   1350  CG   MET B  46      13.153 -15.586  11.187
ATOM   1351  SD   MET B  46      12.977 -15.188   9.473
ATOM   1352  CE   MET B  46      13.566 -16.690   8.775
ATOM   1353  N    ILE B  47      11.933 -11.379  13.030
ATOM   1354  H    ILE B  47      12.327 -10.991  12.196
ATOM   1355  CA   ILE B  47      10.971 -10.568  13.797
ATOM   1356  C    ILE B  47       9.761 -10.233  12.962
ATOM   1357  O    ILE B  47       9.819 -10.048  11.731
ATOM   1358  CB   ILE B  47      11.608  -9.294  14.385
ATOM   1359  CG1  ILE B  47      12.345  -8.459  13.318
ATOM   1360  CG2  ILE B  47      12.542  -9.638  15.494
ATOM   1361  CD1  ILE B  47      12.789  -7.123  13.851
ATOM   1362  N    GLY B  48       8.557 -10.136  13.558
ATOM   1363  H    GLY B  48       8.484 -10.249  14.549
ATOM   1364  CA   GLY B  48       7.365  -9.872  12.800
ATOM   1365  C    GLY B  48       6.826  -8.512  13.141
ATOM   1366  O    GLY B  48       7.136  -7.832  14.149
ATOM   1367  N    GLY B  49       5.940  -8.027  12.306
ATOM   1368  H    GLY B  49       5.668  -8.562  11.506
ATOM   1369  CA   GLY B  49       5.336  -6.745  12.493
ATOM   1370  C    GLY B  49       4.082  -6.786  11.674
ATOM   1371  O    GLY B  49       3.561  -7.847  11.273
ATOM   1372  N    ILE B  50       3.531  -5.634  11.315
ATOM   1373  H    ILE B  50       4.015  -4.777  11.492
ATOM   1374  CA   ILE B  50       2.247  -5.573  10.673
ATOM   1375  C    ILE B  50       2.118  -6.456   9.420
ATOM   1376  O    ILE B  50       1.175  -7.253   9.215
ATOM   1377  CB   ILE B  50       1.982  -4.071  10.391
ATOM   1378  CG1  ILE B  50       1.005  -3.539  11.396
ATOM   1379  CG2  ILE B  50       1.610  -3.739   8.922
ATOM   1380  CD1  ILE B  50      -0.391  -4.077  11.252
ATOM   1381  N    GLY B  51       3.113  -6.410   8.519
ATOM   1382  H    GLY B  51       3.957  -5.920   8.737
ATOM   1383  CA   GLY B  51       2.926  -7.075   7.259
ATOM   1384  C    GLY B  51       3.671  -8.391   7.077
ATOM   1385  O    GLY B  51       3.716  -8.945   5.973
ATOM   1386  N    GLY B  52       4.296  -8.982   8.116
ATOM   1387  H    GLY B  52       4.227  -8.580   9.029
ATOM   1388  CA   GLY B  52       5.053 -10.190   7.874
ATOM   1389  C    GLY B  52       6.334 -10.178   8.678
ATOM   1390  O    GLY B  52       6.519  -9.421   9.657
ATOM   1391  N    PHE B  53       7.325 -11.015   8.343
ATOM   1392  H    PHE B  53       7.227 -11.603   7.540
ATOM   1393  CA   PHE B  53       8.542 -11.096   9.110
ATOM   1394  C    PHE B  53       9.727 -10.584   8.315
ATOM   1395  O    PHE B  53       9.780 -10.618   7.075
ATOM   1396  CB   PHE B  53       8.804 -12.555   9.542
ATOM   1397  CG   PHE B  53       7.850 -13.023  10.592
ATOM   1398  CD1  PHE B  53       6.513 -13.277  10.279
```

FIG. 11Y

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1399 | CD2 | PHE | B | 53 | 8.279 | -13.192 | 11.918 |
| ATOM | 1400 | CE1 | PHE | B | 53 | 5.620 | -13.697 | 11.253 |
| ATOM | 1401 | CE2 | PHE | B | 53 | 7.382 | -13.615 | 12.903 |
| ATOM | 1402 | CZ | PHE | B | 53 | 6.052 | -13.868 | 12.574 |
| ATOM | 1403 | N | ILE | B | 54 | 10.758 | -10.126 | 8.985 |
| ATOM | 1404 | H | ILE | B | 54 | 10.665 | -9.922 | 9.960 |
| ATOM | 1405 | CA | ILE | B | 54 | 12.029 | -9.910 | 8.338 |
| ATOM | 1406 | C | ILE | B | 54 | 13.089 | -10.648 | 9.134 |
| ATOM | 1407 | O | ILE | B | 54 | 12.952 | -11.006 | 10.325 |
| ATOM | 1408 | CB | ILE | B | 54 | 12.390 | -8.444 | 8.236 |
| ATOM | 1409 | CG1 | ILE | B | 54 | 12.386 | -7.775 | 9.611 |
| ATOM | 1410 | CG2 | ILE | B | 54 | 11.460 | -7.770 | 7.218 |
| ATOM | 1411 | CD1 | ILE | B | 54 | 13.113 | -6.438 | 9.590 |
| ATOM | 1412 | N | LYS | B | 55 | 14.272 | -10.852 | 8.523 |
| ATOM | 1413 | H | LYS | B | 55 | 14.383 | -10.599 | 7.562 |
| ATOM | 1414 | CA | LYS | B | 55 | 15.403 | -11.431 | 9.216 |
| ATOM | 1415 | C | LYS | B | 55 | 16.274 | -10.324 | 9.732 |
| ATOM | 1416 | O | LYS | B | 55 | 16.620 | -9.328 | 9.047 |
| ATOM | 1417 | CB | LYS | B | 55 | 16.222 | -12.237 | 8.245 |
| ATOM | 1418 | CG | LYS | B | 55 | 15.638 | -13.596 | 8.063 |
| ATOM | 1419 | CD | LYS | B | 55 | 16.299 | -14.348 | 6.953 |
| ATOM | 1420 | CE | LYS | B | 55 | 15.311 | -14.520 | 5.813 |
| ATOM | 1421 | NZ | LYS | B | 55 | 15.757 | -15.577 | 4.897 |
| ATOM | 1422 | 1HZ | LYS | B | 55 | 15.095 | -15.676 | 4.154 |
| ATOM | 1423 | 3HZ | LYS | B | 55 | 15.830 | -16.441 | 5.395 |
| ATOM | 1424 | 2HZ | LYS | B | 55 | 16.650 | -15.334 | 4.518 |
| ATOM | 1425 | N | VAL | B | 56 | 16.880 | -10.547 | 10.910 |
| ATOM | 1426 | H | VAL | B | 56 | 16.741 | -11.418 | 11.382 |
| ATOM | 1427 | CA | VAL | B | 56 | 17.732 | -9.578 | 11.534 |
| ATOM | 1428 | C | VAL | B | 56 | 18.884 | -10.304 | 12.184 |
| ATOM | 1429 | O | VAL | B | 56 | 18.884 | -11.539 | 12.367 |
| ATOM | 1430 | CB | VAL | B | 56 | 16.912 | -8.819 | 12.609 |
| ATOM | 1431 | CG1 | VAL | B | 56 | 15.865 | -7.943 | 11.921 |
| ATOM | 1432 | CG2 | VAL | B | 56 | 16.215 | -9.788 | 13.599 |
| ATOM | 1433 | N | ARG | B | 57 | 19.958 | -9.593 | 12.591 |
| ATOM | 1434 | H | ARG | B | 57 | 20.030 | -8.624 | 12.353 |
| ATOM | 1435 | CA | ARG | B | 57 | 21.050 | -10.193 | 13.386 |
| ATOM | 1436 | C | ARG | B | 57 | 20.963 | -9.608 | 14.804 |
| ATOM | 1437 | O | ARG | B | 57 | 20.814 | -8.395 | 15.053 |
| ATOM | 1438 | CB | ARG | B | 57 | 22.426 | -9.873 | 12.817 |
| ATOM | 1439 | CG | ARG | B | 57 | 22.664 | -10.437 | 11.439 |
| ATOM | 1440 | CD | ARG | B | 57 | 24.012 | -10.065 | 10.899 |
| ATOM | 1441 | NE | ARG | B | 57 | 24.280 | -10.697 | 9.617 |
| ATOM | 1442 | HE | ARG | B | 57 | 23.592 | -11.323 | 9.250 |
| ATOM | 1443 | CZ | ARG | B | 57 | 25.392 | -10.478 | 8.921 |
| ATOM | 1444 | NH1 | ARG | B | 57 | 26.337 | -9.650 | 9.353 |
| ATOM | 1445 | 2HH1 | ARG | B | 57 | 26.223 | -9.171 | 10.224 |
| ATOM | 1446 | 1HH1 | ARG | B | 57 | 27.163 | -9.505 | 8.808 |
| ATOM | 1447 | NH2 | ARG | B | 57 | 25.561 | -11.104 | 7.760 |
| ATOM | 1448 | 1HH2 | ARG | B | 57 | 26.392 | -10.950 | 7.225 |
| ATOM | 1449 | 2HH2 | ARG | B | 57 | 24.857 | -11.729 | 7.422 |
| ATOM | 1450 | N | GLN | B | 58 | 20.997 | -10.489 | 15.832 |
| ATOM | 1451 | H | GLN | B | 58 | 21.176 | -11.456 | 15.650 |
| ATOM | 1452 | CA | GLN | B | 58 | 20.780 | -10.072 | 17.206 |
| ATOM | 1453 | C | GLN | B | 58 | 22.108 | -9.886 | 17.882 |
| ATOM | 1454 | O | GLN | B | 58 | 22.918 | -10.815 | 18.038 |

FIG. 11Z

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1455 | CB | GLN | B | 58 | 20.051 | -11.190 | 17.932 |
| ATOM | 1456 | CG | GLN | B | 58 | 19.765 | -10.845 | 19.366 |
| ATOM | 1457 | CD | GLN | B | 58 | 19.179 | -12.003 | 20.112 |
| ATOM | 1458 | OE1 | GLN | B | 58 | 19.712 | -12.472 | 21.101 |
| ATOM | 1459 | NE2 | GLN | B | 58 | 18.055 | -12.476 | 19.623 |
| ATOM | 1460 | 1HE2 | GLN | B | 58 | 17.598 | -13.249 | 20.063 |
| ATOM | 1461 | 2HE2 | GLN | B | 58 | 17.647 | -12.066 | 18.807 |
| ATOM | 1462 | N | TYR | B | 59 | 22.416 | -8.692 | 18.422 |
| ATOM | 1463 | H | TYR | B | 59 | 21.788 | -7.921 | 18.311 |
| ATOM | 1464 | CA | TYR | B | 59 | 23.631 | -8.486 | 19.161 |
| ATOM | 1465 | C | TYR | B | 59 | 23.244 | -8.290 | 20.607 |
| ATOM | 1466 | O | TYR | B | 59 | 22.178 | -7.728 | 20.927 |
| ATOM | 1467 | CB | TYR | B | 59 | 24.387 | -7.241 | 18.653 |
| ATOM | 1468 | CG | TYR | B | 59 | 24.271 | -7.075 | 17.149 |
| ATOM | 1469 | CD1 | TYR | B | 59 | 23.045 | -7.242 | 16.494 |
| ATOM | 1470 | CD2 | TYR | B | 59 | 25.385 | -6.753 | 16.374 |
| ATOM | 1471 | CE1 | TYR | B | 59 | 22.939 | -7.093 | 15.112 |
| ATOM | 1472 | CE2 | TYR | B | 59 | 25.291 | -6.603 | 14.995 |
| ATOM | 1473 | CZ | TYR | B | 59 | 24.068 | -6.774 | 14.365 |
| ATOM | 1474 | OH | TYR | B | 59 | 24.018 | -6.620 | 13.010 |
| ATOM | 1475 | HH | TYR | B | 59 | 24.926 | -6.394 | 12.658 |
| ATOM | 1476 | N | ASP | B | 60 | 24.010 | -8.785 | 21.596 |
| ATOM | 1477 | H | ASP | B | 60 | 24.852 | -9.276 | 21.372 |
| ATOM | 1478 | CA | ASP | B | 60 | 23.644 | -8.624 | 22.992 |
| ATOM | 1479 | C | ASP | B | 60 | 24.556 | -7.595 | 23.615 |
| ATOM | 1480 | O | ASP | B | 60 | 25.654 | -7.261 | 23.125 |
| ATOM | 1481 | CB | ASP | B | 60 | 23.789 | -9.920 | 23.777 |
| ATOM | 1482 | CG | ASP | B | 60 | 22.803 | -10.960 | 23.332 |
| ATOM | 1483 | OD1 | ASP | B | 60 | 21.619 | -10.634 | 23.032 |
| ATOM | 1484 | OD2 | ASP | B | 60 | 23.208 | -12.126 | 23.273 |
| ATOM | 1485 | N | GLN | B | 61 | 24.156 | -7.022 | 24.774 |
| ATOM | 1486 | H | GLN | B | 61 | 23.252 | -7.234 | 25.146 |
| ATOM | 1487 | CA | GLN | B | 61 | 25.011 | -6.086 | 25.519 |
| ATOM | 1488 | C | GLN | B | 61 | 25.411 | -4.866 | 24.746 |
| ATOM | 1489 | O | GLN | B | 61 | 26.560 | -4.382 | 24.832 |
| ATOM | 1490 | CB | GLN | B | 61 | 26.269 | -6.763 | 26.028 |
| ATOM | 1491 | CG | GLN | B | 61 | 26.020 | -8.038 | 26.753 |
| ATOM | 1492 | CD | GLN | B | 61 | 25.714 | -7.766 | 28.185 |
| ATOM | 1493 | OE1 | GLN | B | 61 | 24.572 | -7.455 | 28.548 |
| ATOM | 1494 | NE2 | GLN | B | 61 | 26.744 | -7.844 | 29.014 |
| ATOM | 1495 | 1HE2 | GLN | B | 61 | 26.620 | -7.675 | 29.992 |
| ATOM | 1496 | 2HE2 | GLN | B | 61 | 27.654 | -8.073 | 28.669 |
| ATOM | 1497 | N | ILE | B | 62 | 24.539 | -4.257 | 23.933 |
| ATOM | 1498 | H | ILE | B | 62 | 23.628 | -4.648 | 23.801 |
| ATOM | 1499 | CA | ILE | B | 62 | 24.878 | -3.047 | 23.238 |
| ATOM | 1500 | C | ILE | B | 62 | 24.571 | -1.885 | 24.144 |
| ATOM | 1501 | O | ILE | B | 62 | 23.515 | -1.819 | 24.819 |
| ATOM | 1502 | CB | ILE | B | 62 | 24.097 | -2.922 | 21.912 |
| ATOM | 1503 | CG1 | ILE | B | 62 | 24.310 | -4.170 | 21.094 |
| ATOM | 1504 | CG2 | ILE | B | 62 | 24.568 | -1.709 | 21.067 |
| ATOM | 1505 | CD1 | ILE | B | 62 | 25.794 | -4.479 | 20.878 |
| ATOM | 1506 | N | LEU | B | 63 | 25.485 | -0.912 | 24.304 |
| ATOM | 1507 | H | LEU | B | 63 | 26.403 | -1.028 | 23.926 |
| ATOM | 1508 | CA | LEU | B | 63 | 25.192 | 0.322 | 25.015 |
| ATOM | 1509 | C | LEU | B | 63 | 24.630 | 1.296 | 24.030 |
| ATOM | 1510 | O | LEU | B | 63 | 25.239 | 1.658 | 22.995 |

FIG. 11aa

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1511 | CB | LEU | B | 63 | 26.436 | 0.970 | 25.590 |
| ATOM | 1512 | CG | LEU | B | 63 | 26.186 | 2.358 | 26.226 |
| ATOM | 1513 | CD1 | LEU | B | 63 | 25.486 | 2.261 | 27.576 |
| ATOM | 1514 | CD2 | LEU | B | 63 | 27.468 | 3.162 | 26.382 |
| ATOM | 1515 | N | ILE | B | 64 | 23.492 | 1.946 | 24.358 |
| ATOM | 1516 | H | ILE | B | 64 | 22.958 | 1.643 | 25.148 |
| ATOM | 1517 | CA | ILE | B | 64 | 23.003 | 3.068 | 23.617 |
| ATOM | 1518 | C | ILE | B | 64 | 22.872 | 4.194 | 24.612 |
| ATOM | 1519 | O | ILE | B | 64 | 22.915 | 4.007 | 25.846 |
| ATOM | 1520 | CB | ILE | B | 64 | 21.634 | 2.701 | 22.989 |
| ATOM | 1521 | CG1 | ILE | B | 64 | 21.825 | 1.521 | 22.029 |
| ATOM | 1522 | CG2 | ILE | B | 64 | 20.982 | 3.894 | 22.246 |
| ATOM | 1523 | CD1 | ILE | B | 64 | 20.593 | 1.096 | 21.260 |
| ATOM | 1524 | N | GLU | B | 65 | 22.803 | 5.460 | 24.172 |
| ATOM | 1525 | H | GLU | B | 65 | 23.013 | 5.664 | 23.216 |
| ATOM | 1526 | CA | GLU | B | 65 | 22.432 | 6.551 | 25.037 |
| ATOM | 1527 | C | GLU | B | 65 | 21.242 | 7.194 | 24.373 |
| ATOM | 1528 | O | GLU | B | 65 | 21.312 | 7.729 | 23.257 |
| ATOM | 1529 | CB | GLU | B | 65 | 23.497 | 7.615 | 25.131 |
| ATOM | 1530 | CG | GLU | B | 65 | 24.787 | 7.196 | 25.761 |
| ATOM | 1531 | CD | GLU | B | 65 | 25.694 | 8.385 | 26.076 |
| ATOM | 1532 | OE1 | GLU | B | 65 | 25.170 | 9.510 | 26.311 |
| ATOM | 1533 | OE2 | GLU | B | 65 | 26.938 | 8.200 | 26.092 |
| ATOM | 1534 | N | ILE | B | 66 | 20.078 | 7.240 | 25.035 |
| ATOM | 1535 | H | ILE | B | 66 | 20.010 | 6.835 | 25.947 |
| ATOM | 1536 | CA | ILE | B | 66 | 18.907 | 7.865 | 24.462 |
| ATOM | 1537 | C | ILE | B | 66 | 18.777 | 9.195 | 25.145 |
| ATOM | 1538 | O | ILE | B | 66 | 18.591 | 9.303 | 26.379 |
| ATOM | 1539 | CB | ILE | B | 66 | 17.713 | 6.995 | 24.790 |
| ATOM | 1540 | CG1 | ILE | B | 66 | 17.916 | 5.583 | 24.335 |
| ATOM | 1541 | CG2 | ILE | B | 66 | 16.405 | 7.544 | 24.177 |
| ATOM | 1542 | CD1 | ILE | B | 66 | 16.888 | 4.677 | 24.884 |
| ATOM | 1543 | N | CYS | B | 67 | 18.965 | 10.325 | 24.437 |
| ATOM | 1544 | H | CYS | B | 67 | 19.201 | 10.268 | 23.467 |
| ATOM | 1545 | CA | CYS | B | 67 | 18.833 | 11.663 | 25.049 |
| ATOM | 1546 | C | CYS | B | 67 | 19.637 | 11.781 | 26.319 |
| ATOM | 1547 | O | CYS | B | 67 | 19.235 | 12.400 | 27.328 |
| ATOM | 1548 | CB | CYS | B | 67 | 17.387 | 12.023 | 25.319 |
| ATOM | 1549 | SG | CYS | B | 67 | 16.407 | 12.259 | 23.821 |
| ATOM | 1550 | N | GLY | B | 68 | 20.830 | 11.180 | 26.383 |
| ATOM | 1551 | H | GLY | B | 68 | 21.158 | 10.646 | 25.604 |
| ATOM | 1552 | CA | GLY | B | 68 | 21.654 | 11.288 | 27.558 |
| ATOM | 1553 | C | GLY | B | 68 | 21.464 | 10.185 | 28.584 |
| ATOM | 1554 | O | GLY | B | 68 | 22.174 | 10.128 | 29.606 |
| ATOM | 1555 | N | HIS | B | 69 | 20.513 | 9.255 | 28.425 |
| ATOM | 1556 | H | HIS | B | 69 | 19.924 | 9.282 | 27.618 |
| ATOM | 1557 | CA | HIS | B | 69 | 20.304 | 8.199 | 29.391 |
| ATOM | 1558 | C | HIS | B | 69 | 20.861 | 6.936 | 28.811 |
| ATOM | 1559 | O | HIS | B | 69 | 20.589 | 6.560 | 27.647 |
| ATOM | 1560 | CB | HIS | B | 69 | 18.832 | 7.992 | 29.654 |
| ATOM | 1561 | CG | HIS | B | 69 | 18.175 | 9.203 | 30.223 |
| ATOM | 1562 | ND1 | HIS | B | 69 | 17.504 | 9.195 | 31.435 |
| ATOM | 1563 | HD1 | HIS | B | 69 | 17.383 | 8.402 | 32.032 |
| ATOM | 1564 | CD2 | HIS | B | 69 | 18.122 | 10.470 | 29.729 |
| ATOM | 1565 | CE1 | HIS | B | 69 | 17.070 | 10.429 | 31.626 |
| ATOM | 1566 | NE2 | HIS | B | 69 | 17.410 | 11.240 | 30.635 |

FIG. 11bb

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1567 | N | LYS | B | 70 | 21.751 | 6.217 | 29.499 |
| ATOM | 1568 | H | LYS | B | 70 | 22.025 | 6.512 | 30.414 |
| ATOM | 1569 | CA | LYS | B | 70 | 22.326 | 5.020 | 28.945 |
| ATOM | 1570 | C | LYS | B | 70 | 21.386 | 3.854 | 29.145 |
| ATOM | 1571 | O | LYS | B | 70 | 20.627 | 3.725 | 30.120 |
| ATOM | 1572 | CB | LYS | B | 70 | 23.613 | 4.678 | 29.663 |
| ATOM | 1573 | CG | LYS | B | 70 | 24.694 | 5.655 | 29.379 |
| ATOM | 1574 | CD | LYS | B | 70 | 25.739 | 5.524 | 30.444 |
| ATOM | 1575 | CE | LYS | B | 70 | 27.048 | 6.090 | 30.011 |
| ATOM | 1576 | NZ | LYS | B | 70 | 26.948 | 7.548 | 30.000 |
| ATOM | 1577 | 1HZ | LYS | B | 70 | 27.821 | 7.940 | 29.711 |
| ATOM | 1578 | 3HZ | LYS | B | 70 | 26.725 | 7.874 | 30.919 |
| ATOM | 1579 | 2HZ | LYS | B | 70 | 26.230 | 7.828 | 29.363 |
| ATOM | 1580 | N | ALA | B | 71 | 21.512 | 2.849 | 28.284 |
| ATOM | 1581 | H | ALA | B | 71 | 22.141 | 2.934 | 27.512 |
| ATOM | 1582 | CA | ALA | B | 71 | 20.762 | 1.630 | 28.432 |
| ATOM | 1583 | C | ALA | B | 71 | 21.629 | 0.576 | 27.805 |
| ATOM | 1584 | O | ALA | B | 71 | 22.463 | 0.830 | 26.912 |
| ATOM | 1585 | CB | ALA | B | 71 | 19.452 | 1.726 | 27.737 |
| ATOM | 1586 | N | ILE | B | 72 | 21.547 | -0.681 | 28.237 |
| ATOM | 1587 | H | ILE | B | 72 | 20.864 | -0.925 | 28.926 |
| ATOM | 1588 | CA | ILE | B | 72 | 22.424 | -1.698 | 27.730 |
| ATOM | 1589 | C | ILE | B | 72 | 21.615 | -2.938 | 27.462 |
| ATOM | 1590 | O | ILE | B | 72 | 20.909 | -3.490 | 28.330 |
| ATOM | 1591 | CB | ILE | B | 72 | 23.524 | -1.999 | 28.737 |
| ATOM | 1592 | CG1 | ILE | B | 72 | 24.322 | -0.735 | 29.090 |
| ATOM | 1593 | CG2 | ILE | B | 72 | 24.442 | -3.037 | 28.153 |
| ATOM | 1594 | CD1 | ILE | B | 72 | 25.374 | -1.012 | 30.163 |
| ATOM | 1595 | N | GLY | B | 73 | 21.609 | -3.446 | 26.235 |
| ATOM | 1596 | H | GLY | B | 73 | 22.204 | -3.054 | 25.534 |
| ATOM | 1597 | CA | GLY | B | 73 | 20.707 | -4.545 | 26.062 |
| ATOM | 1598 | C | GLY | B | 73 | 20.828 | -5.084 | 24.663 |
| ATOM | 1599 | O | GLY | B | 73 | 21.754 | -4.831 | 23.863 |
| ATOM | 1600 | N | THR | B | 74 | 19.856 | -5.905 | 24.271 |
| ATOM | 1601 | H | THR | B | 74 | 19.086 | -6.088 | 24.882 |
| ATOM | 1602 | CA | THR | B | 74 | 19.869 | -6.548 | 22.988 |
| ATOM | 1603 | C | THR | B | 74 | 19.363 | -5.590 | 21.931 |
| ATOM | 1604 | O | THR | B | 74 | 18.338 | -4.870 | 22.053 |
| ATOM | 1605 | CB | THR | B | 74 | 19.011 | -7.801 | 23.074 |
| ATOM | 1606 | OG1 | THR | B | 74 | 19.611 | -8.683 | 24.013 |
| ATOM | 1607 | HG1 | THR | B | 74 | 19.068 | -9.519 | 24.092 |
| ATOM | 1608 | CG2 | THR | B | 74 | 18.817 | -8.496 | 21.705 |
| ATOM | 1609 | N | VAL | B | 75 | 20.028 | -5.620 | 20.762 |
| ATOM | 1610 | H | VAL | B | 75 | 20.835 | -6.203 | 20.666 |
| ATOM | 1611 | CA | VAL | B | 75 | 19.630 | -4.837 | 19.611 |
| ATOM | 1612 | C | VAL | B | 75 | 19.600 | -5.771 | 18.426 |
| ATOM | 1613 | O | VAL | B | 75 | 20.444 | -6.673 | 18.230 |
| ATOM | 1614 | CB | VAL | B | 75 | 20.667 | -3.712 | 19.395 |
| ATOM | 1615 | CG1 | VAL | B | 75 | 20.473 | -3.002 | 18.046 |
| ATOM | 1616 | CG2 | VAL | B | 75 | 20.679 | -2.708 | 20.567 |
| ATOM | 1617 | N | LEU | B | 76 | 18.557 | -5.647 | 17.565 |
| ATOM | 1618 | H | LEU | B | 76 | 17.822 | -5.000 | 17.767 |
| ATOM | 1619 | CA | LEU | B | 76 | 18.444 | -6.427 | 16.324 |
| ATOM | 1620 | C | LEU | B | 76 | 18.736 | -5.487 | 15.144 |
| ATOM | 1621 | O | LEU | B | 76 | 18.239 | -4.343 | 15.040 |
| ATOM | 1622 | CB | LEU | B | 76 | 17.028 | -7.021 | 16.158 |

FIG. 11cc

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1623 | CG | LEU | B | 76 | 16.427 | -7.612 | 17.449 |
| ATOM | 1624 | CD1 | LEU | B | 76 | 14.992 | -8.075 | 17.263 |
| ATOM | 1625 | CD2 | LEU | B | 76 | 17.266 | -8.758 | 18.019 |
| ATOM | 1626 | N | VAL | B | 77 | 19.607 | -5.900 | 14.222 |
| ATOM | 1627 | H | VAL | B | 77 | 19.985 | -6.824 | 14.276 |
| ATOM | 1628 | CA | VAL | B | 77 | 20.027 | -5.042 | 13.133 |
| ATOM | 1629 | C | VAL | B | 77 | 19.570 | -5.662 | 11.842 |
| ATOM | 1630 | O | VAL | B | 77 | 19.678 | -6.883 | 11.598 |
| ATOM | 1631 | CB | VAL | B | 77 | 21.563 | -4.905 | 13.191 |
| ATOM | 1632 | CG1 | VAL | B | 77 | 22.129 | -4.202 | 11.944 |
| ATOM | 1633 | CG2 | VAL | B | 77 | 22.030 | -4.166 | 14.470 |
| ATOM | 1634 | N | GLY | B | 78 | 18.978 | -4.915 | 10.943 |
| ATOM | 1635 | H | GLY | B | 78 | 18.841 | -3.941 | 11.121 |
| ATOM | 1636 | CA | GLY | B | 78 | 18.523 | -5.475 | 9.705 |
| ATOM | 1637 | C | GLY | B | 78 | 18.019 | -4.338 | 8.874 |
| ATOM | 1638 | O | GLY | B | 78 | 18.130 | -3.142 | 9.223 |
| ATOM | 1639 | N | PRO | B | 79 | 17.408 | -4.596 | 7.722 |
| ATOM | 1640 | CA | PRO | B | 79 | 16.954 | -3.535 | 6.834 |
| ATOM | 1641 | C | PRO | B | 79 | 15.635 | -2.872 | 7.280 |
| ATOM | 1642 | O | PRO | B | 79 | 14.609 | -2.877 | 6.565 |
| ATOM | 1643 | CB | PRO | B | 79 | 16.804 | -4.274 | 5.492 |
| ATOM | 1644 | CG | PRO | B | 79 | 16.463 | -5.712 | 5.881 |
| ATOM | 1645 | CD | PRO | B | 79 | 17.159 | -5.959 | 7.189 |
| ATOM | 1646 | N | THR | B | 80 | 15.574 | -2.247 | 8.458 |
| ATOM | 1647 | H | THR | B | 80 | 16.374 | -2.242 | 9.058 |
| ATOM | 1648 | CA | THR | B | 80 | 14.364 | -1.583 | 8.865 |
| ATOM | 1649 | C | THR | B | 80 | 14.312 | -0.189 | 8.228 |
| ATOM | 1650 | O | THR | B | 80 | 15.349 | 0.471 | 8.001 |
| ATOM | 1651 | CB | THR | B | 80 | 14.250 | -1.512 | 10.410 |
| ATOM | 1652 | OG1 | THR | B | 80 | 13.079 | -0.802 | 10.806 |
| ATOM | 1653 | HG1 | THR | B | 80 | 13.022 | -0.766 | 11.804 |
| ATOM | 1654 | CG2 | THR | B | 80 | 15.519 | -0.901 | 11.062 |
| ATOM | 1655 | N | PRO | B | 81 | 13.137 | 0.354 | 7.885 |
| ATOM | 1656 | CA | PRO | B | 81 | 13.036 | 1.747 | 7.379 |
| ATOM | 1657 | C | PRO | B | 81 | 13.363 | 2.732 | 8.484 |
| ATOM | 1658 | O | PRO | B | 81 | 13.791 | 3.880 | 8.250 |
| ATOM | 1659 | CB | PRO | B | 81 | 11.548 | 1.912 | 6.982 |
| ATOM | 1660 | CG | PRO | B | 81 | 10.819 | 0.674 | 7.488 |
| ATOM | 1661 | CD | PRO | B | 81 | 11.854 | -0.387 | 7.797 |
| ATOM | 1662 | N | VAL | B | 82 | 13.197 | 2.368 | 9.772 |
| ATOM | 1663 | H | VAL | B | 82 | 12.940 | 1.427 | 9.992 |
| ATOM | 1664 | CA | VAL | B | 82 | 13.380 | 3.306 | 10.885 |
| ATOM | 1665 | C | VAL | B | 82 | 14.160 | 2.668 | 12.010 |
| ATOM | 1666 | O | VAL | B | 82 | 14.045 | 1.465 | 12.293 |
| ATOM | 1667 | CB | VAL | B | 82 | 11.996 | 3.695 | 11.431 |
| ATOM | 1668 | CG1 | VAL | B | 82 | 12.055 | 4.961 | 12.269 |
| ATOM | 1669 | CG2 | VAL | B | 82 | 10.958 | 3.857 | 10.318 |
| ATOM | 1670 | N | ASN | B | 83 | 14.963 | 3.422 | 12.775 |
| ATOM | 1671 | H | ASN | B | 83 | 15.147 | 4.370 | 12.516 |
| ATOM | 1672 | CA | ASN | B | 83 | 15.550 | 2.846 | 13.967 |
| ATOM | 1673 | C | ASN | B | 83 | 14.481 | 2.874 | 15.022 |
| ATOM | 1674 | O | ASN | B | 83 | 13.814 | 3.903 | 15.294 |
| ATOM | 1675 | CB | ASN | B | 83 | 16.743 | 3.639 | 14.472 |
| ATOM | 1676 | CG | ASN | B | 83 | 17.935 | 3.574 | 13.570 |
| ATOM | 1677 | OD1 | ASN | B | 83 | 18.409 | 2.511 | 13.167 |
| ATOM | 1678 | ND2 | ASN | B | 83 | 18.439 | 4.735 | 13.238 |

FIG. 1 ldd

```
ATOM   1679 2HD2 ASN B  83      19.237   4.786  12.638
ATOM   1680 1HD2 ASN B  83      18.030   5.580  13.582
ATOM   1681  N   ILE B  84      14.225   1.749  15.711
ATOM   1682  H   ILE B  84      14.791   0.938  15.564
ATOM   1683  CA  ILE B  84      13.154   1.658  16.667
ATOM   1684  C   ILE B  84      13.740   1.317  18.020
ATOM   1685  O   ILE B  84      14.428   0.300- 18.223
ATOM   1686  CB  ILE B  84      12.214   0.517  16.260
ATOM   1687  CG1 ILE B  84      11.656   0.759  14.849
ATOM   1688  CG2 ILE B  84      11.128   0.247  17.315
ATOM   1689  CD1 ILE B  84      10.770  -0.359  14.291
ATOM   1690  N   ILE B  85      13.483   2.157  19.051
ATOM   1691  H   ILE B  85      13.028   3.030  18.877
ATOM   1692  CA  ILE B  85      13.846   1.834  20.408
ATOM   1693  C   ILE B  85      12.596   1.254  21.085
ATOM   1694  O   ILE B  85      11.536   1.903  21.267
ATOM   1695  CB  ILE B  85      14.308   3.115  21.137
ATOM   1696  CG1 ILE B  85      15.447   3.826  20.395
ATOM   1697  CG2 ILE B  85      14.673   2.840  22.589
ATOM   1698  CD1 ILE B  85      16.730   3.053  20.263
ATOM   1699  N   GLY B  86      12.617  -0.052  21.422
ATOM   1700  H   GLY B  86      13.439  -0.595  21.251
ATOM   1701  CA  GLY B  86      11.481  -0.702  22.028
ATOM   1702  C   GLY B  86      11.557  -0.748  23.538
ATOM   1703  O   GLY B  86      12.412  -0.165  24.238
ATOM   1704  N   ARG B  87      10.614  -1.489  24.149
ATOM   1705  H   ARG B  87      10.012  -2.072  23.604
ATOM   1706  CA  ARG B  87      10.442  -1.468  25.584
ATOM   1707  C   ARG B  87      11.627  -2.021  26.326
ATOM   1708  O   ARG B  87      11.911  -1.666  27.495
ATOM   1709  CB  ARG B  87       9.200  -2.271  25.949
ATOM   1710  CG  ARG B  87       7.951  -1.960  25.161
ATOM   1711  CD  ARG B  87       6.956  -3.074  25.219
ATOM   1712  NE  ARG B  87       5.906  -2.933  24.205
ATOM   1713  HE  ARG B  87       5.790  -2.039  23.772
ATOM   1714  CZ  ARG B  87       5.119  -3.953  23.856
ATOM   1715  NH1 ARG B  87       5.252  -5.161  24.396
ATOM   1716 2HH1 ARG B  87       5.958  -5.326  25.085
ATOM   1717 1HH1 ARG B  87       4.646  -5.905  24.113
ATOM   1718  NH2 ARG B  87       4.180  -3.751  22.939
ATOM   1719 1HH2 ARG B  87       3.580  -4.502  22.664
ATOM   1720 2HH2 ARG B  87       4.073  -2.848  22.524
ATOM   1721  N   ASN B  88      12.413  -2.937  25.731
ATOM   1722  H   ASN B  88      12.206  -3.237  24.800
ATOM   1723  CA  ASN B  88      13.582  -3.519  26.415
ATOM   1724  C   ASN B  88      14.532  -2.429  26.821
ATOM   1725  O   ASN B  88      15.214  -2.516  27.863
ATOM   1726  CB  ASN B  88      14.285  -4.605  25.559
ATOM   1727  CG  ASN B  88      15.063  -4.031  24.358
ATOM   1728  OD1 ASN B  88      14.515  -3.245  23.612
ATOM   1729  ND2 ASN B  88      16.333  -4.445  24.180
ATOM   1730 2HD2 ASN B  88      16.875  -4.099  23.414
ATOM   1731 1HD2 ASN B  88      16.744  -5.102  24.812
ATOM   1732  N   LEU B  89      14.695  -1.328  26.061
ATOM   1733  H   LEU B  89      14.192  -1.240  25.201
ATOM   1734  CA  LEU B  89      15.597  -0.234  26.452
```

FIG. 11ee

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1735 | C | LEU | B | 89 | 14.797 | 0.937 | 27.053 |
| ATOM | 1736 | O | LEU | B | 89 | 15.293 | 1.734 | 27.879 |
| ATOM | 1737 | CB | LEU | B | 89 | 16.421 | 0.232 | 25.236 |
| ATOM | 1738 | CG | LEU | B | 89 | 17.400 | -0.754 | 24.567 |
| ATOM | 1739 | CD1 | LEU | B | 89 | 18.215 | 0.002 | 23.573 |
| ATOM | 1740 | CD2 | LEU | B | 89 | 18.352 | -1.458 | 25.570 |
| ATOM | 1741 | N | LEU | B | 90 | 13.511 | 1.114 | 26.705 |
| ATOM | 1742 | H | LEU | B | 90 | 13.082 | 0.486 | 26.056 |
| ATOM | 1743 | CA | LEU | B | 90 | 12.698 | 2.221 | 27.257 |
| ATOM | 1744 | C | LEU | B | 90 | 12.537 | 2.060 | 28.751 |
| ATOM | 1745 | O | LEU | B | 90 | 12.575 | 3.033 | 29.533 |
| ATOM | 1746 | CB | LEU | B | 90 | 11.311 | 2.258 | 26.628 |
| ATOM | 1747 | CG | LEU | B | 90 | 11.232 | 2.730 | 25.168 |
| ATOM | 1748 | CD1 | LEU | B | 90 | 9.808 | 2.744 | 24.642 |
| ATOM | 1749 | CD2 | LEU | B | 90 | 11.831 | 4.105 | 24.982 |
| ATOM | 1750 | N | THR | B | 91 | 12.315 | 0.843 | 29.271 |
| ATOM | 1751 | H | THR | B | 91 | 12.218 | 0.055 | 28.663 |
| ATOM | 1752 | CA | THR | B | 91 | 12.210 | 0.634 | 30.699 |
| ATOM | 1753 | C | THR | B | 91 | 13.537 | 1.028 | 31.375 |
| ATOM | 1754 | O | THR | B | 91 | 13.575 | 1.525 | 32.518 |
| ATOM | 1755 | CB | THR | B | 91 | 11.893 | -0.843 | 31.028 |
| ATOM | 1756 | OG1 | THR | B | 91 | 12.919 | -1.676 | 30.504 |
| ATOM | 1757 | HG1 | THR | B | 91 | 12.722 | -2.634 | 30.713 |
| ATOM | 1758 | CG2 | THR | B | 91 | 10.599 | -1.285 | 30.418 |
| ATOM | 1759 | N | GLN | B | 92 | 14.705 | 0.852 | 30.732 |
| ATOM | 1760 | H | GLN | B | 92 | 14.707 | 0.497 | 29.797 |
| ATOM | 1761 | CA | GLN | B | 92 | 15.920 | 1.190 | 31.433 |
| ATOM | 1762 | C | GLN | B | 92 | 16.088 | 2.660 | 31.633 |
| ATOM | 1763 | O | GLN | B | 92 | 16.807 | 3.137 | 32.527 |
| ATOM | 1764 | CB | GLN | B | 92 | 17.127 | 0.680 | 30.682 |
| ATOM | 1765 | CG | GLN | B | 92 | 17.076 | -0.805 | 30.517 |
| ATOM | 1766 | CD | GLN | B | 92 | 18.336 | -1.314 | 29.900 |
| ATOM | 1767 | OE1 | GLN | B | 92 | 19.394 | -0.720 | 30.059 |
| ATOM | 1768 | NE2 | GLN | B | 92 | 18.221 | -2.411 | 29.195 |
| ATOM | 1769 | 1HE2 | GLN | B | 92 | 19.022 | -2.813 | 28.751 |
| ATOM | 1770 | 2HE2 | GLN | B | 92 | 17.331 | -2.856 | 29.095 |
| ATOM | 1771 | N | ILE | B | 93 | 15.538 | 3.512 | 30.746 |
| ATOM | 1772 | H | ILE | B | 93 | 15.016 | 3.153 | 29.972 |
| ATOM | 1773 | CA | ILE | B | 93 | 15.693 | 4.937 | 30.899 |
| ATOM | 1774 | C | ILE | B | 93 | 14.522 | 5.549 | 31.698 |
| ATOM | 1775 | O | ILE | B | 93 | 14.438 | 6.773 | 31.940 |
| ATOM | 1776 | CB | ILE | B | 93 | 15.981 | 5.657 | 29.548 |
| ATOM | 1777 | CG1 | ILE | B | 93 | 14.746 | 5.718 | 28.619 |
| ATOM | 1778 | CG2 | ILE | B | 93 | 17.223 | 5.060 | 28.874 |
| ATOM | 1779 | CD1 | ILE | B | 93 | 14.946 | 6.734 | 27.488 |
| ATOM | 1780 | N | GLY | B | 94 | 13.617 | 4.731 | 32.263 |
| ATOM | 1781 | H | GLY | B | 94 | 13.639 | 3.752 | 32.060 |
| ATOM | 1782 | CA | GLY | B | 94 | 12.594 | 5.224 | 33.170 |
| ATOM | 1783 | C | GLY | B | 94 | 11.443 | 5.846 | 32.432 |
| ATOM | 1784 | O | GLY | B | 94 | 10.766 | 6.803 | 32.878 |
| ATOM | 1785 | N | CYS | B | 95 | 11.134 | 5.354 | 31.225 |
| ATOM | 1786 | H | CYS | B | 95 | 11.603 | 4.538 | 30.888 |
| ATOM | 1787 | CA | CYS | B | 95 | 10.134 | 5.969 | 30.381 |
| ATOM | 1788 | C | CYS | B | 95 | 8.750 | 5.512 | 30.764 |
| ATOM | 1789 | O | CYS | B | 95 | 8.478 | 4.309 | 31.006 |
| ATOM | 1790 | CB | CYS | B | 95 | 10.456 | 5.643 | 28.922 |

FIG. 1 Iff

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ATOM | 1791 | SG | CYS | B | 95 | 9.426 | 6.512 | 27.764 |
| ATOM | 1792 | N | THR | B | 96 | 7.778 | 6.444 | 30.764 |
| ATOM | 1793 | H | THR | B | 96 | 8.014 | 7.401 | 30.539 |
| ATOM | 1794 | CA | THR | B | 96 | 6.379 | 6.163 | 31.108 |
| ATOM | 1795 | C | THR | B | 96 | 5.390 | 6.970 | 30.254 |
| ATOM | 1796 | O | THR | B | 96 | 5.567 | 8.171 | 30.066 |
| ATOM | 1797 | CB | THR | B | 96 | 6.111 | 6.439 | 32.604 |
| ATOM | 1798 | OG1 | THR | B | 96 | 6.341 | 7.794 | 32.938 |
| ATOM | 1799 | HG1 | THR | B | 96 | 6.111 | 7.924 | 33.861 |
| ATOM | 1800 | CG2 | THR | B | 96 | 6.938 | 5.566 | 33.554 |
| ATOM | 1801 | N | LEU | B | 97 | 4.302 | 6.321 | 29.809 |
| ATOM | 1802 | H | LEU | B | 97 | 4.216 | 5.332 | 29.997 |
| ATOM | 1803 | CA | LEU | B | 97 | 3.127 | 6.986 | 29.238 |
| ATOM | 1804 | C | LEU | B | 97 | 2.336 | 7.681 | 30.358 |
| ATOM | 1805 | O | LEU | B | 97 | 2.350 | 7.221 | 31.499 |
| ATOM | 1806 | CB | LEU | B | 97 | 2.226 | 5.958 | 28.532 |
| ATOM | 1807 | CG | LEU | B | 97 | 2.860 | 5.279 | 27.300 |
| ATOM | 1808 | CD1 | LEU | B | 97 | 2.101 | 3.986 | 26.957 |
| ATOM | 1809 | CD2 | LEU | B | 97 | 2.842 | 6.216 | 26.085 |
| ATOM | 1810 | N | ASN | B | 98 | 1.637 | 8.777 | 30.024 |
| ATOM | 1811 | H | ASN | B | 98 | 1.662 | 9.086 | 29.063 |
| ATOM | 1812 | CA | ASN | B | 98 | 0.906 | 9.631 | 30.960 |
| ATOM | 1813 | C | ASN | B | 98 | -0.251 | 10.321 | 30.231 |
| ATOM | 1814 | O | ASN | B | 98 | -0.032 | 11.303 | 29.522 |
| ATOM | 1815 | CB | ASN | B | 98 | 1.845 | 10.678 | 31.587 |
| ATOM | 1816 | CG | ASN | B | 98 | 2.783 | 10.077 | 32.634 |
| ATOM | 1817 | OD1 | ASN | B | 98 | 3.926 | 9.739 | 32.335 |
| ATOM | 1818 | ND2 | ASN | B | 98 | 2.297 | 9.942 | 33.870 |
| ATOM | 1819 | 2HD2 | ASN | B | 98 | 2.877 | 9.551 | 34.599 |
| ATOM | 1820 | 1HD2 | ASN | B | 98 | 1.351 | 10.229 | 34.074 |
| ATOM | 1821 | N | LEU | B | 99 | -1.476 | 9.808 | 30.426 |
| ATOM | 1822 | H | LEU | B | 99 | -1.568 | 9.010 | 31.037 |
| ATOM | 1823 | CA | LEU | B | 99 | -2.709 | 10.288 | 29.797 |
| ATOM | 1824 | C | LEU | B | 99 | -3.816 | 10.589 | 30.815 |
| ATOM | 1825 | O | LEU | B | 99 | -3.630 | 10.272 | 32.011 |
| ATOM | 1826 | CB | LEU | B | 99 | -3.146 | 9.340 | 28.657 |
| ATOM | 1827 | CG | LEU | B | 99 | -3.714 | 7.932 | 28.941 |
| ATOM | 1828 | CD1 | LEU | B | 99 | -2.767 | 7.057 | 29.774 |
| ATOM | 1829 | CD2 | LEU | B | 99 | -5.134 | 7.943 | 29.528 |
| ATOM | 1830 | OXT | LEU | B | 99 | -4.842 | 11.156 | 30.376 |
| TER | | | | | | | | |

FIG. 11gg

SEQUENCE LISTING

<110> Kalyanaraman Ramnarayan
Edward T. Maggio
P. Patrick Hess

<120> Use of Computationally Derived Protein Structures of Genetic Polymorphisms in Pharmacogenomics for Drug Design and Clinical Applications

<130> 24737-1906PC

<140> Unassigned
<141> 2000-11-10

<150> 09/438,566
<151> 1999-11-10

<150> 24737-1906B
<151> 2000-11-01

<160> 118

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified Hepatitis C Virus NS3 Protease Inhibitor Peptide

<221> ACETYLATION
<222> 1

<221> MOD_RES
<222> 2
<223> D-glutamic acid

<221> MOD_RES
<222> 5
<223> beta-cyclohexylalanine

<300>
<301> Ingallinella, P., Altamura, S., Bianchi, E., Talia
<302> Potent Peptide Inhibitors Of Human Hepatitis C Vir
<303> Biochemistry
<304> 37
<305> 25
<306> 8906-8914
<307> 1998-06-23

<400> 1
Asp Xaa Leu Ile Xaa Cys
1 5

<210> 2
<211> 6

<212> PRT
<213> Artificial Sequence

<220>
<223> Modified Hepatitis C Virus NS3 Protease Inhibitor Peptide

<221> ACETYLATION
<222> 1

<221> MOD_RES
<222> 5
<223> beta-cyclohexylalanine

<300>
<301> Ingallinella, P., Altamura, S., Bianchi, E., Talia
<302> Potent Peptide Inhibitors Of Human Hepatitis C Vir
<303> Biochemistry
<304> 37
<305> 25
<306> 8906-8914
<307> 1998-06-23

<400> 2
Asp Glu Leu Ile Xaa Cys
1 5

<210> 3
<211> 1045
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> Protease

<221> CDS
<222> (298)...(1045)
<223> Portion of Reverse Transcriptase

<400> 3
cct cag atc act ctt tgg caa cga ccc cty gtc aca ata aag ata ggg 48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
1 5 10 15

ggc caa cta aaa gaa gct yta tta gat aca gga gca gat gat aca gta 96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

tta gaa gaa atg agt tta cca ggg aaa tgg aaa cca aaa atg ata ggg 144
Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
35 40 45

gga att gga ggt ttt atc aaa gta aga cag tat gat caa ata ctc ata 192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
50 55 60

gaa atc tgt gga cat aaa gct ata ggc aca gta tta gta gga cct aca 240

```
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttg ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Leu Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gag aat cca tac aat act cca ata ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga aca caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cac ccc gca ggg tta aaa cag aaa aaa tca gta aca ata ctg      624
Ile Pro His Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Ile Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa ggc ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Gly Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aga aat aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Arg Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aac gtg ctc cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aga aty tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Xaa Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gga cat cta tta aag tgg gga ttt acc aca cca gac aaa aaa cat      960
```

```
Arg Gly His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat      1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata aag ttg cca gaa aaa g                    1045
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys
            340                 345
```

<210> 4
<211> 1046
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1046)
<223> Portion of HIV Reverse Transcriptase

<400> 4

```
cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aag ata gga       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

gtt gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg      144
Val Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gag caa ata gcc gta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Glu Gln Ile Ala Val
    50                  55                  60

gaa aty tgt gga cat aga gct atg ggt aca gta tta gta gga cct aca      240
Glu Xaa Cys Gly His Arg Ala Met Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa ttg gaa aag gaa ggg      432
```

```
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gag gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cca gca ggg tta aaa aag aat aaa tca ata aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Ile Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta tgt gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Cys Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aac aat gag act cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga ttc acc      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Phe Thr
                245                 250                 255

agc ata ttc caa tgt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ser Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gag ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa tat ctg tgg aag tgg gga ttt tgc aca cca gaa caa aar cat      960
Arg Gln Tyr Leu Trp Lys Trp Gly Phe Cys Thr Pro Glu Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cct ttc ctt tgg atg ggt tat gaa ctc cat ccc gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa cct ata gtg ctg cca gac aaa ga                  1046
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Asp Lys
            340                 345
```

<210> 5
<211> 1104
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1104)
<223> Portion of HIV Reverse Transcriptase

<400> 5

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag rta ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Xaa Gly
 1               5                  10                  15

ggg caa cta agg gaa gct cta tta gat aca gga gca gat gat aca ata        96
Gly Gln Leu Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Ile
            20                  25                  30

ata gaa gac ata act ttg cca gga aga tgg aca cca aaa atg ata ggg       144
Ile Glu Asp Ile Thr Leu Pro Gly Arg Trp Thr Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aga cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgc gga cat aaa rtt ata agt aca gta ttg gta gga cct aca       240
Glu Ile Cys Gly His Lys Xaa Ile Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cca ata aac ata gtt gga aga aat ctg atg act cag att ggt tgc act       288
Pro Ile Asn Ile Val Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gtc aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aag gca tta gta gaa att tgt mca gaa ctg gaa atg gat gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Xaa Glu Leu Glu Met Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat ccg tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aac aaa aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aag aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
```

```
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta tgt gaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Cys Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gaa atg gtc atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Met Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gat gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Asp Glu Leu
    290                 295                 300

aga caa tat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln Tyr Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aca gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Thr Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

<210> 6
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 6

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
```

```
1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gat atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agg cag tat gat caa ata ctc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga agg aat ctg ttg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gag gag     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga     576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aag tca gta aca gta ctg     624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg     672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt act ata cct agt ata aac aat gag aca cca ggg     720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca     768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
```

```
                245                 250                 255

gca ata ttc caa agt agc atg ata aaa atc tta gag cct ttc aga aaa      816
Ala Ile Phe Gln Ser Ser Met Ile Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg gtc atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cac aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aag aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata aag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 7
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 7

```
cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aar ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gag gaa atn aat tta cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Xaa Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
```

```
          35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctt gta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Val
     50                  55                  60

gaa aty tgt gga cat aar gct ata ggt aca gta tta gta gga cct aca      240
Glu Xaa Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

ccc gtc aac ata att gga aga aat ctg ttg act caa att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

ccg gga atg gat ggc ccc aaa gtt aaa cat ggc cct ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys His Gly Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aag cct tta gtt gaa att tgt aca gaa atg gga aaa gaa ggg      432
Lys Ile Lys Pro Leu Val Glu Ile Cys Thr Glu Met Gly Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac tty tgg gaa gtc caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc tca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ser Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc ttg gat gaa gac tta gag      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Leu Glu
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
```

```
        275                 280                 285

gga tca gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg ggg tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Gly Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
            325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca aca aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Thr Lys Asp Ser Trp Thr
        340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aac tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
    355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
370
```

<210> 8
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 8

```
cct cag atc act ctt tgg caa cga ccc cty gtc aca gta aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa ata aag gaa gct yta tta gat aca gga gca gat gat aca gta       96
Gly Gln Ile Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ata ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga caa aaa gct ata agt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly Gln Lys Ala Ile Ser Thr Val Leu Val Gly Pro Thr
```

```
65                  70                  75                  80

cct gtc aat ata att gga aga aat ctg atg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa ggc agt aac aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Gly Ser Asn Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg cta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa ggg tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

agc tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa cta      912
Ser Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga tta acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His
```

```
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gar aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 9
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 9

```
cct cag atc act ctt tgg caa cga ccc cty gtc aaa gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Lys Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata atw gga aga aat ctg ttg act cag att ggt tgc act     288
Pro Val Asn Ile Xaa Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
```

```
                100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gag atg gag aag gaa ggg      432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gay ttc tgg gaa gtt car tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aag atc tta gar cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tcw gac tta gaa ata ggg caa cat aga ata aaa ata gag gaa ctg      912
Gly Xaa Asp Leu Glu Ile Gly Gln His Arg Ile Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg tta agg tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggk tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gay agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
```

```
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

atc tac cca ggg                                                  1116
Ile Tyr Pro Gly
    370


<210> 10
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 10
cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa ata aag gaa gct yta tta gat aca gga gca gat gat aca gta      96
Gly Gln Ile Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atw ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Xaa Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga caa aaa gct ata agt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly Gln Lys Ala Ile Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aat ata att gga aga aat ctg atg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa taa aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys  *  Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
```

```
        130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
    145                 150                 155

gcc ata aag aaa aaa ggc agt aac aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Gly Ser Asn Arg Trp Arg Lys Leu Val Asp Phe
160                 165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
                180                 185                 190

ata cca cat ccc gca ggg cta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
            195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
        210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
    225                 230                 235

att aga tat cag tac aat gtg ctt ccm caa gga tgg aaa ggg tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Xaa Gln Gly Trp Lys Gly Ser Pro
240                 245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
                260                 265                 270

caa aat cca gac wtr gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Xaa Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
            275                 280                 285

agc tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa cta      912
Ser Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
        290                 295                 300

aga caa cat ctg ttg agg tgg gga tta acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His
    305                 310                 315

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
320                 325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gag aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
                340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
            355                 360                 365

att tac cca ggg                                                   1116
Ile Tyr Pro Gly
```

370

<210> 11
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 11

```
cct cag atc act ctt tgg caa cga ccc aty gtt aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aaa raa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Xaa Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata gtg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Val
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aga cag tat gat cag gta ccc ata     192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gag atc tgt ggg cat aaa att ata ggt aca gta tta ata gga cct acc     240
Glu Ile Cys Gly His Lys Ile Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gcc aac gta att gga aga aat ctg atg act cag ctt ggt tgc act     288
Pro Ala Asn Val Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt yct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Xaa Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt gca gaa ctg gag aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Ala Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aga att ggg cct gaa aat cca tac aat act cca ata ttt     480
Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aag aaa aac agt act agg tgg aga aaa tta gta gat ttc     528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
```

```
                  165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggg gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa gat agc atg aca aaa atc tta gat ccc ttt aga aag      816
Ala Ile Phe Gln Asp Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

aaa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Lys Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac yta gaa ata gag cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Xaa Glu Ile Glu Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa tat ctg tta aag tgg gga ttt ttc aca cca gag caa aaa cat      960
Arg Glu Tyr Leu Leu Lys Trp Gly Phe Phe Thr Pro Glu Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggc tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aac tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 12
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 12

```
cct caa atc act ctt tgg car cga ccc tta gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gcc cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

cta gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agg cag tat gat car ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gag atc tgc ggg tat aaa gct gtg ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly Tyr Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act caa att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac ggt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Gly Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
```

```
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat caa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Gln Asp Phe Arg
    210                 215                 220

aag tat act gca ttc act ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg gtt atc tat caa tat atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

ggc tct gac tta gaa aya ggg cag cat aga rca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Xaa Gly Gln His Arg Xaa Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata atg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Met Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag cta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 13
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 13

```
cct cag atc act ctt tgg caa cga ccc aty gtc aac ata aag gta ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Asn Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta arg gaa gct cta ata gat aca gga gca gat gat aca gta        96
Gly Gln Leu Xaa Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata gat ttg cca gga aga tgg aga cca aga atg ata ggg       144
Leu Glu Asp Ile Asp Leu Pro Gly Arg Trp Arg Pro Arg Met Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aag cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Val Lys Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa ata tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct acg       240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggg tgc act       288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aaa       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aag ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aag aaa aac agt act aga tgg aga aaa tta gta gat ttt       528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgt gaa gtg caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Cys Glu Val Gln Leu Gly
            180                 185                 190

ata ccg cat ccc gca ggg tta ara aag aaa aga tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Xaa Lys Lys Arg Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gcc ttt acc ata cct agt ata aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc cta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca grc ata gtt atc gtt caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Xaa Ile Val Ile Val Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

ggg tct gac tta gaa ata ggg caa cat aga gca aaa ata gag gag ttg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg ttg agg tgg gga tty ttc aca cca gac gaa aaa cat      960
Arg Glu His Leu Leu Arg Trp Gly Phe Phe Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cac cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg acc gta cag cct ata aat ttg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Asn Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac tca ggg                                                    1116
Ile Tyr Ser Gly
    370
```

<210> 14
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

```
<400> 14
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa gta agg gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Val Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
tta gaa gaa atg aat ttg cca gga aaa tgg aag cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
          35                  40                  45

gga att ggg ggc ttt atc aaa gta aga cag tat gat caa ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
      50                  55                  60

gaa atc tgt gga cat aaa gct ata ggg aca gtg tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt cct att agt cct att gaa act gtg cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttg ata gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aac agt act aga tgg agg aaa cta gta gac ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca gga tta aaa aag aga aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Arg Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aar gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tac act gca ttc acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctg cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

tca ata ttc caa agt agy atg aca aaa atc tta gag cct ttt aga aag      816
Ser Ile Phe Gln Ser Xaa Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gac ata gat atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Asp Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata rag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Xaa Glu Leu
    290                 295                 300

aga gag cat ctg cta aag tgg gga ttt acc aca cca gac raa aaa cat      960
Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Xaa Lys His
305                 310                 315                 320

car aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctt cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa cat ata gag cta cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln His Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 15
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 15

```
cct caa atc act ctt tgg car cga ccc ctc gtt gca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta kaa gaa atg gat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Xaa Glu Met Asp Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta tcc wta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Ser Xaa
    50                  55                  60
```

```
gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tat aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa ttg gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gar gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat cca gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Pro Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agy agc atg ata aga aty tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Xaa Ser Met Ile Arg Xaa Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata gaa cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300
```

```
aga caa cat ctg tta agg tgg gga ttt acc aca cca gay aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggk tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag ttr gtg gga aaa ttr aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Xaa Val Gly Lys Xaa Asn Trp Ala Ser Gln
        355                 360                 365

att tac tca ggg                                                    1116
Ile Tyr Ser Gly
    370


<210> 16
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 16
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gag gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg act ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Met Thr Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta rta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Xaa Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aar gat ggt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Gly Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aag tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt act ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa atc tta gag cct ttt aga aag      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga rca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Xaa Lys Ile Glu Glu Leu
    290                 295                 300

agg caa cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cca gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag cct ata gtg ctg cca caa aaa gac agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Gln Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                     1116
Ile Tyr Pro Gly
    370
```

<210> 17
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 17

```
cct caa atc act ctt tgg caa cga ccc aty gtc aca ata aag gta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gcc cta ata gat aca gga gca gat gat aca gtg       96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ttg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag rta ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct gta ggt tca gtg tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Val Gly Ser Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

cta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu
        115                 120                 125
```

```
aaa ata gaa gca tta gta gaa atc tgt gca gaa ctg gaa gag gca ggg      432
Lys Ile Glu Ala Leu Val Glu Ile Cys Ala Glu Leu Glu Glu Ala Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aar aag aac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aac aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca att ccc tta gat aag gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt aca ata cct agy ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Xaa Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cma cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Xaa Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc cag tgt agc atg aca aaa atc tta gat cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg car cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa yat ctg tgg aag tgg gga ttt tac aca cca gag aat aaa cat      960
Arg Gln Xaa Leu Trp Lys Trp Gly Phe Tyr Thr Pro Glu Asn Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc cwt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Xaa Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tat gcn ggg                                                        1116
Ile Tyr Ala Gly
    370
```

<210> 18
<211> 1117
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1117)
<223> Portion of HIV Reverse Transcriptase

<400> 18

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg         48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg car cta aag gaa gct cta tta gat aca gga gca gat gat aca gta         96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

gta gaa gaa atg aat tta tca gga agg tgg aaa cca aaa atg ata ggg        144
Val Glu Glu Met Asn Leu Ser Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga saa tat gaa cag ata cct gta        192
Gly Ile Gly Gly Phe Ile Lys Val Arg Xaa Tyr Glu Gln Ile Pro Val
    50                  55                  60

gaa att tgt gga cat aaa gct gta ggt aca gta tta gtg gga cct aca        240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act        288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt ccc att gaa act gta cca gta aaa ttg aag        336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc ccg aga gtt aaa caa tgg cca ttg aca gaa gaa        384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg        432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt        480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aaa gac agt aat aaa tgg agg aaa tta gtg gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Asn Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta ggg      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccy tca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Xaa Ser Gly Leu Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tac ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att agr tat cag tac aat gtg ctg cca cag gga tgg aaa gga tca cca      768
Ile Xaa Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga gaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Glu
            260                 265                 270

caa aat aca gac ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Thr Asp Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa gtr gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Xaa Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga yta acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Xaa Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc cgt tgg atg ggk tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Arg Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtr caa cct ata gag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Xaa Gln Pro Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aaa gtt agt ggg aaa att aaa ttg ggc aag tca     1104
Val Asn Asp Ile Gln Lys Val Ser Gly Lys Ile Lys Leu Gly Lys Ser
        355                 360                 365

gat tta ccc agg g                                                  1117
Asp Leu Pro Arg
    370
```

<210> 19
<211> 1116

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 19

```
cct cag atc act ctt tgg caa cga ccc cty gtc aca gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta acg gaa gct yta ttg gat aca gga gca gat aat aca gta      96
Gly Gln Leu Thr Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asn Thr Val
            20                  25                  30

tta gaa gaa atg agt ttr cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Ser Xaa Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gta gta ggt aca gta tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Val Val Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga gat ctg ttg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gtg aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aar gac agt act aaa tgg aga aaa ttr gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Xaa Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga     576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc tca ggg tta aaa aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Ser Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gac gtg ggt gat gca tat ttc tca gtt ccc cta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttc acc ata cct agt gta aac aat gag act cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctg cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cac aat cca aac ata gtt atc tat caa tac gtg gat gat tta tat gta      864
His Asn Pro Asn Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa gta gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Val Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg ggg ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 20
<211> 1117
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1117)
<223> Portion of HIV Reverse Transcriptase

<400> 20

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata gga        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata aat ttg cca ggg aaa tgg aaa cca aaa atg ata ggg       144
Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat caa ata cca gta       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
    50                  55                  60

gaa att tgt gga cat aaa gct gta ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac gta att gga aga aat ctg atg act cag att ggc tgc act       288
Pro Val Asn Val Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aaa gtt aaa caa tgg cca tta aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgc aca gaa ttg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att gga cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca ata ccc tta gat gaa gaa ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Glu Glu Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt cca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Pro Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa tgt agt atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

gaa aat cca gat ata gtt atc tac caa tac atg gat gac tta tat gta      864
Glu Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa tat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln Tyr Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag caa gaa cct cca ttc cgt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Gln Glu Pro Pro Phe Arg Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag ttt agt ggg aaa att gaa ttg ggc aag tca     1104
Val Asn Asp Ile Gln Lys Phe Ser Gly Lys Ile Glu Leu Gly Lys Ser
        355                 360                 365

gat tta tgc agg g                                                  1117
Asp Leu Cys Arg
    370
```

<210> 21
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

```
<400> 21
cct cag atc act ctt tgg caa cga mcc gtt gtc wca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Xaa Val Val Xaa Ile Lys Ile Gly
 1               5                  10                  15
```

```
ggg caa cta aaa gaa gct cta tta gay aca ggg gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
                20                  25                  30

tta gaa gac atg cat ttg cca ggt aga tgg aaa cca aaa atg ata gtg       144
Leu Glu Asp Met His Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Val
            35                  40                  45

gga att ggg ggt ttt gtc aaa gta aga cag tat gat cag ata cct gta       192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
        50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cca gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act       288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttc ccc atc agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa att aga caa tgg cca tta aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Ile Arg Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aat agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt atg aac aat gag aca cca gga       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Met Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca atg gga tgg aaa gga tca cca       768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Met Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttc caa agt agt atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtc atc tat caa tac atg gat gat tta tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tcg gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ttg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aga tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtt aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 22
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 22

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag gta gga      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gag gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata gat ttg cca gga agr tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Ile Asp Leu Pro Gly Xaa Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
     50                  55                  60

gaa ata tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cgg att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Arg Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttt      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtg caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata ccg cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aar gay ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gcc ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc cta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285
```

```
ggg tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg ttg agg tgg gga ttt acc acc cca gac aaa aaa cat      960
Arg Glu His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg acc gtr cag cct ata gag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Xaa Gln Pro Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370


<210> 23
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 23
cct cag atc act ctt tgg caa cga ccc ata gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta ata gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata aat ttg cca gga aga tgg aaa cca aaa tta ata ggg      144
Leu Glu Asp Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aga gtg aaa cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Val Arg Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa att tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80
```

```
cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta aca gaa atc tgt wca gag atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Thr Glu Ile Cys Xaa Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aac act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcy ata cac aag aaa aat agt aat aga tgg aga aaa gta gta gat ttc      528
Xaa Ile His Lys Lys Asn Ser Asn Arg Trp Arg Lys Val Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca gga tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aag gat ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gcg ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tac cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aga atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata aak gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Xaa Glu Leu
    290                 295                 300

aga saa cat ctg ttg agg tgg gga ttt ttc aca cca gac caa aaa cat      960
Arg Xaa His Leu Leu Arg Trp Gly Phe Phe Thr Pro Asp Gln Lys His
305                 310                 315                 320
```

```
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat      1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aar gac agt tgg acw      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Xaa
            340                 345                 350

gty aat gac ata cag aaa tta gtk gga aaa ttg aat tgg gca agt caa      1104
Xaa Asn Asp Ile Gln Lys Leu Xaa Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                     1116
Ile Tyr Pro Gly
    370
```

<210> 24
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 24

```
cct cag atc act ctt tgg caa cga ccc ata gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta ata gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata aat ttg cca gga aga tgg aaa cca aaa tta ata ggg      144
Leu Glu Asp Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aga gtg aaa cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Val Arg Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa att tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110
```

```
cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta aca gaa atc tgt wca gag atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Thr Glu Ile Cys Xaa Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aac act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcy ata cac aag aaa aat agt aat aga tgg aga aaa gta gta gat ttc      528
Xaa Ile His Lys Lys Asn Ser Asn Arg Trp Arg Lys Val Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca gga tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aag gat ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gcg ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tac cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aga atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata aak gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Xaa Glu Leu
    290                 295                 300

aga saa cat ctg ttg agg tgg gga ttt ttc aca cca gac caa aaa cat      960
Arg Xaa His Leu Leu Arg Trp Gly Phe Phe Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aar gac agt tgg acw     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Xaa
            340                 345                 350
```

```
gty aat gac ata cag aaa tta gtk gga aaa ttg aat tgg gca agt caa     1104
Xaa Asn Asp Ile Gln Lys Leu Xaa Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 25
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 25

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aaa ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta cta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agt ttg cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta tcc atg     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Ser Met
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga tct aca     240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Ser Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ytg ttg act cag ctt ggg tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Xaa Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gar ggg     432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140
```

```
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aag tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gat ttc tgg gaa rtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Xaa Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta caa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Gln Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtc ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tat agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Tyr Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tac caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctc tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtt cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 26
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 26

```
cct cag atc act ctt tgg caa cga ccc atc gtc gaa ata aag gta ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Glu Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat tta cca gga aga tgg aaa cca aga atg ata ggg     144
Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Arg Met Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aga cag tat gat cag gta cct atc     192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata agt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Val Ile Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg atg act cag att ggt tgc act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aaa     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ytg gaa gag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Xaa Glu Glu Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aag aaa nnn agt ggt aga tgg aga aaa ata gta gat ttt     528
Ala Ile Lys Lys Lys Xaa Ser Gly Arg Trp Arg Lys Ile Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aac aag tca gta aca att ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Ile Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aag gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aat aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat cag tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata ggg gag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Glu His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga car cat ctg tta arg tgg gga ttt ttc aca cca gaa caa aaa cat      960
Arg Gln His Leu Leu Xaa Trp Gly Phe Phe Thr Pro Glu Gln Lys His
305                 310                 315                 320

cag aaa gaa cct ccm ttc cak tgg atg ggt tat gaa ctc cay cct gat     1008
Gln Lys Glu Pro Xaa Phe Xaa Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cas cct ata gtg ctg cca gaa aaa gat agc tgg act     1056
Lys Trp Thr Val Xaa Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 27
<211> 1113
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 27

```
cct cag atc act ctt tgg caa cga ccc atc gtc gaa ata aag gta ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Glu Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat tta cca gga aga tgg aaa cca aga atg ata ggg     144
Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Arg Met Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aga cag tat gat cag gta cct atc     192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata agt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Val Ile Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg atg act cag att ggt tgc act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aaa     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ytg gaa gag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Xaa Glu Glu Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aag aaa agt ggt aga tgg aga aaa ata gta gat ttt aga     528
Ala Ile Lys Lys Lys Ser Gly Arg Trp Arg Lys Ile Val Asp Phe Arg
                165                 170                 175

gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga ata     576
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
            180                 185                 190

cca cat ccc gca ggg tta aaa aag aac aag tca gta aca att ctg gat     624
Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Ile Leu Asp
        195                 200                 205
```

```
gtg ggt gat gca tat ttt tca gtt ccc tta gat aag gaa ttc agg aag      672
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg Lys
    210                 215                 220

tat act gca ttt acc ata cct agt ata aat aat gag aca cca ggg att      720
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
225                 230                 235                 240

aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca gca      768
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
                245                 250                 255

ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa caa      816
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
            260                 265                 270

aat cca gac ata gtt atc tat cag tac gtg gat gat ttg tat gta gga      864
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val Gly
        275                 280                 285

tct gat tta gaa ata ggg gag cat aga aca aaa ata gag gaa ctg aga      912
Ser Asp Leu Glu Ile Gly Glu His Arg Thr Lys Ile Glu Glu Leu Arg
    290                 295                 300

car cat ctg tta arg tgg gga ttt ttc aca cca gaa caa aaa cat cag      960
Gln His Leu Leu Xaa Trp Gly Phe Phe Thr Pro Glu Gln Lys His Gln
305                 310                 315                 320

aaa gaa cct ccm ttc cak tgg atg ggt tat gaa ctc cay cct gat aaa     1008
Lys Glu Pro Xaa Phe Xaa Trp Met Gly Tyr Glu Leu His Pro Asp Lys
                325                 330                 335

tgg aca gta cas cct ata gtg ctg cca gaa aaa gat agc tgg act gtc     1056
Trp Thr Val Xaa Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
            340                 345                 350

aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag att     1104
Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
        355                 360                 365

tac cca ggg                                                        1113
Tyr Pro Gly
    370
```

<210> 28
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 28

```
cct caa atc act stt tgg caa cga ccc aty gtc tca ata aag ata ggg      48
Pro Gln Ile Thr Xaa Trp Gln Arg Pro Xaa Val Ser Ile Lys Ile Gly
 1               5                  10                  15

ggg caa ata aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Ile Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aag cca aaa atg ata gtg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Val
        35                  40                  45

gga att gga ggt ttt agc aaa gta aga caa tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ser Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgc gga cgt aaa gtt gta ggt tca gta tta ata gga cct aca     240
Glu Ile Cys Gly Arg Lys Val Val Gly Ser Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag ctt ggc tgt act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct atk gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Xaa Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gag     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa ttg gaa gaa gma gga     432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Leu Glu Glu Xaa Gly
    130                 135                 140

aaa att aca aaa att ggg cct gaa aat ccg tac aat act cca ata ttt     480
Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aar aac agt act aaa tgg aga aaa tta gta gac ttc     528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga     576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg     624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat aaa gac ttc agg     672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aar tat act gca ttt acc ata cct agt acg aat aat gag aca cca ggg     720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg ctt gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gag gag cta      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gka cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Xaa Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 29
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

```
<400> 29
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta ata gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ata ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
          35                  40                  45

gga att gga ggt ctt gtc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Leu Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
      50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gtw tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Xaa Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca ggg atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gag aag gag gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act agg tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aac aaa tca gca aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Ala Thr Val Leu
        195                 200                 205

gat gtg ggc gat gca tat ttt tca gtt ccc tta gac aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acy ata cct agt ata aac aat gaa aca cca ggg      720
Lys Tyr Thr Ala Phe Xaa Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

tar ata tca gtg tac aat gtr ctt cca caa gga tgg aaa gga tca cma      768
Xaa Ile Ser Val Tyr Asn Xaa Leu Pro Gln Gly Trp Lys Gly Ser Xaa
                245                 250                 255

gca ata ttc maa agt agc atg aca aga atc tta gag cct ttt aga aaa      816
Ala Ile Phe Xaa Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gaa ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa gta gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Val Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt ttc aca cca gac caa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Phe Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac gcn ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 30
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 30

```
cct caa atc act ctt tgg caa cga ccc cty gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct yta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agc tta cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Ser Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggk ttt atc aaa gtg agm cag tat gat cag ata ctc ata      192
Gly Ile Gly Xaa Phe Ile Lys Val Xaa Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60
```

```
gaa aty tgt gga cat aaa gct ata ggt aca gtr tta ata gga cct aca      240
Glu Xaa Cys Gly His Lys Ala Ile Gly Thr Xaa Leu Ile Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa ttg aaa      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtc aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggr      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Xaa
    130                 135                 140

aaa att aca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aag aaa aac agt gat aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Asp Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cca gca ggg tta aaa cag aaa aag tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gta ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

gtt aga tat cag tac aat gta ctc cca cag gga tgg aaa gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aaa atc tta gag cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300
```

```
aga caa cat ctg ttg agg tgg gga ttc tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gta ggg aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gga                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 31
<211> 1117
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 31

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa tta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

cta gaa gac gtg cat ttg cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Val His Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat gag gta ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Glu Val Pro Ile
    50                  55                  60

gaa ctc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Leu Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

ccc gtc aac ata att gga aga aat ctg wtg act caa ctt ggg tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Xaa Thr Gln Leu Gly Cys Thr
                85                  90                  95
```

```
cta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aga gtt ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Arg Val Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gyc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cay ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctr      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Xaa
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tac cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gat cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

caa aac cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tcy gac tta gaa ata gga cag cat agr rca aaa ata gaa gaa ctg      912
Gly Xaa Asp Leu Glu Ile Gly Gln His Xaa Xaa Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aag aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

car aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gtg cag cct ata gtg ctg cca gaa aag gac agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ant aca gaa gtt agt ggg aaa att gaa ttg ggc aag tca      1104
Val Asn Asp Xaa Thr Glu Val Ser Gly Lys Ile Glu Leu Gly Lys Ser
        355                 360                 365

gat tta tgc agg g                                                   1117
Asp Leu Cys Arg
    370
```

<210> 32
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 32

```
cct caa atc act ctt tgg caa cga ccc cty gtc gca ata agg ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Ala Ile Arg Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gcc cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg gag ttg cca gga aga tgg aag cca aaa atg ata ggg      144
Leu Glu Asp Met Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aam cag tat gat cag ata ctt gta      192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Gln Tyr Asp Gln Ile Leu Val
    50                  55                  60

gaa atc tgt gga cat aaa gct gta ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg ttg act cag att ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gag      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125
```

```
aaa ata aaa gca tta gta gaa atc tgt aca gaa ttg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tcc gtg aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttt aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc aya cct sgt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Xaa Pro Xaa Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tcc cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agc agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac wta gtt wtc tat caa twc ata gat gat ctg tat gta      864
Gln Asn Pro Asp Xaa Val Xaa Tyr Gln Xaa Ile Asp Asp Leu Tyr Val
        275                 280                 285

ggc tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg tgg aag tgg ggg ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata atg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Met Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aar tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tac cca ggg                                                          1116
Ile Tyr Pro Gly
    370


<210> 33
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 33
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg          48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta kat aca gga gca gat gat aca gtm          96
Gly Gln Leu Lys Glu Ala Leu Leu Xaa Thr Gly Ala Asp Asp Thr Xaa
            20                  25                  30

tta gaa gac atg act ttg cca gga aga tgg aaa cca aaa atg ata ggg         144
Leu Glu Asp Met Thr Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gag gag ata ccc ata         192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Glu Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca         240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg ttg act cag att ggt tgc act         288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aaa         336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa         384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttw gta gaa att tgt gca gaa ctg gaa aag gaa ggg         432
Lys Ile Lys Ala Xaa Val Glu Ile Cys Ala Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt         480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aaa gac ggt act aaa tgg aga aag gta aca gat ttt      528
Ala Ile Lys Lys Lys Asp Gly Thr Lys Trp Arg Lys Val Thr Asp Phe
                165                 170                 175

aga gaa ctt aat aag agg ach caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Xaa Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc tca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ser Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gcg aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Ala Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg ggt ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat tca ggg                                                   1116
Ile Tyr Ser Gly
    370
```

<210> 34
<211> 1119

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 34

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta ttr gac aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Xaa Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ata ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt att aaa gta aaa cag tat gaa cag ata acc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Gln Ile Thr Ile
    50                  55                  60

gam atc tgt gga cat aaa gct aca ggt aca gta tta gta gga cct aca       240
Xaa Ile Cys Gly His Lys Ala Thr Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac gta att gga aga aat atg atg act cag att ggt tgc act       288
Pro Val Asn Val Ile Gly Arg Asn Met Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aac aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc gca ggg tta cca aag aac aaa tca gta acg gta ctg      624
Ile Pro His Pro Ala Gly Leu Pro Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt cct tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agg tat aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Arg Tyr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

act aga tat cag tac aat gtg ctt cct atg gga tgg aaa gga tca cca      768
Thr Arg Tyr Gln Tyr Asn Val Leu Pro Met Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aga      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Arg
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gac ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gag ata ggg cag cat aga gcg aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg tgg aag tgg ggt ttt tac aca cca gac aaa aaa cat      960
Arg Glu His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc cat tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe His Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg ggr aaa att gaa ttt ggg cga gtc     1104
Val Asn Asp Ile Gln Lys Leu Val Xaa Lys Ile Glu Phe Gly Arg Val
        355                 360                 365

aga ttt amc caa ggg                                                1119
Arg Phe Xaa Gln Gly
    370
```

<210> 35
<211> 1115
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 35

```
cct cag atc act ctt tgg caa cga ccc cty gtc cca ata arg ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Pro Ile Xaa Ile Gly
 1               5                  10                  15

ggg caa tta aag gaa gct cta cta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg aat tta cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Asp Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aar gta aaa cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt ggg cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

cta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att gga cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa gtc caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta tta       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg gga gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtc ata tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

ggg tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cac ttg ttg maa tgg gga ttc acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Xaa Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata kaa ctg cca gaa aaa gac agc tgg ctg     1056
Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Glu Lys Asp Ser Trp Leu
            340                 345                 350

tca atg aca tac aga aat tag tgg gaa agt tga att ggg caa gtc aaa     1104
Ser Met Thr Tyr Arg Asn  *  Trp Glu Ser  *  Ile Gly Gln Val Lys
        355                 360                 365

ttt atg cng gg                                                     1115
Phe Met Xaa
```

<210> 36
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

```
<400> 36
cct cag atc act ctt tgg caa cga cca gtc gtc aca ata aag gta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Ile Lys Val Gly
 1               5                  10                  15
```

```
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
                20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
            35                  40                  45

gga att gga ggt ttt rtc aaa gta aga cag tat gat caa ata ccc ata       192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
        50                  55                  60

gaa atc tgt gga cat aaa gct aca ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Thr Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gyc aac ata att gga aga aat ctg ttg act cag att ggg tgc act       288
Pro Xaa Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ctg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt gca gaa ttg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Ala Glu Leu Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg ccy gaa aat cca tac aay act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Xaa Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aar aac agt act ara tgg aga aaa kta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Xaa Trp Arg Lys Xaa Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg cta aag aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc ttg gat gaa gac ttc aga       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat aca gcc ttt acc tat act ggt tcc aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Tyr Thr Gly Ser Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat car tac aat gtg ctt cca cag gga tgg aaa gga tca cca       768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttc caa agc agc atg aca aaa gtc tta gaa cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Val Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tta agg tgg gga ttt tac aca cca gac gaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Tyr Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gac     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtt aat gac ata cag aaa tta gtg gga aaa ttg aat tgg gcc agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 37
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 37

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aaa ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Asp Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
     50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg aca cag ctt ggt tgt act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta ggg      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca gga tta aaa aag aat aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gat cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

cag aat cca gat ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285
```

```
gga tct gac tta gag ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gca cat ctg ttg aag tgg gga ttt acc acc cca gac aaa aaa cat      960
Arg Ala His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 38
<211> 1117
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1117)
<223> Portion of HIV Reverse Transcriptase

<400> 38

```
cct caa tca ctt ctt tgg caa cga ccc mtc gtc aca ata aag gta ggg      48
Pro Gln Ser Leu Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca ata      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Ile
            20                  25                  30

tta gaa gac aya rat ttg cca ggg aga tgg aaa cca aaa ata ata ggg     144
Leu Glu Asp Xaa Xaa Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aga gta aga cag tat gat cag gta ccc ata     192
Gly Ile Gly Gly Phe Ile Arg Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt gta agt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Val Val Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80
```

```
cct gcc aac ata att gga aga aat ctg atg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt gaa gaa ttg gaa aag gat ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Glu Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca gga tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat gaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

tca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ser Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtc atc tat caa tat atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gag ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg tgg aag tgg ggg ttt tac aca cca gac ara aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Xaa Lys His
305                 310                 315                 320
```

```
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gac      1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tan tsc agg g                                                   1117
Ile Xaa Xaa Arg
    370
```

<210> 39
<211> 1128
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1128)
<223> Portion of HIV Reverse Transcriptase

<400> 39

```
cct cag atc act ctt tgg caa cga cca ttc gtc aca ata aaa ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct ata tta gac aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Ile Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt mtc aaa gta aga cag tat gat cag gta ccc ata      192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt atg agt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Val Met Ser Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg act cag mtt ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Xaa Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gwa cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Xaa Pro Val Lys Leu Lys
            100                 105                 110
```

```
cca ggg atg gac ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt aat aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Asn Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtc caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat raa gat tca gra      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Xaa Asp Ser Xaa
    210                 215                 220

agt aca ctg cat tta cca tac cta gta cgr acc aat gag aca cca ggg      720
Ser Thr Leu His Leu Pro Tyr Leu Val Xaa Thr Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac tta gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata gag cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aag tgg ggg ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc cgt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Arg Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa gcc tat aaa gct gcc aga aaa aga cag ctg gac     1056
Lys Trp Thr Val Gln Ala Tyr Lys Ala Ala Arg Lys Arg Gln Leu Asp
            340                 345                 350
```

```
tgt caa tga cat tac mag aaa gtt agt ggg gaa aat tgg aat ttg ggg      1104
Cys Gln  *  His Tyr Xaa Lys Val Ser Gly Glu Asn Trp Asn Leu Gly
           355                 360                 365

caa ggt cag att tat tgc cag ggg                                     1128
Gln Gly Gln Ile Tyr Cys Gln Gly
       370                 375


<210> 40
<211> 1120
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1120)
<223> Portion of HIV Reverse Transcriptase

<400> 40
cct cag atc act ctt tgg caa cga ccc ctc gtt gca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly
 1               5                  10                  15

gga cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agt ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Ser Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga car tat gat cag ata ccm rta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Xaa Xaa
    50                  55                  60

gaa att tgc gga cat aaa gct gta ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag mtt ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Xaa Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gtg aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140
```

```
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt cct tta gat gaa gac ttc agr      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Xaa
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tcc aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Ser Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc cta gaa cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gat ata gtt atc tat caa tac atg gat gat cta tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg ggg ttt acc acc cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac nat aca aaa gtt agt ggg gaa aat tga att ggg sca agt     1104
Val Asn Asp Xaa Thr Lys Val Ser Gly Glu Asn  *  Ile Gly Xaa Ser
        355                 360                 365

cag att tat tgg agg g                                              1120
Gln Ile Tyr Trp Arg
    370
```

<210> 41
<211> 1059
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1059)
<223> Portion of HIV Reverse Transcriptase

<400> 41

```
cct caa atc act ctt tgg cag cga ccc gtt gtc aca ata aac ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Ile Asn Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gac aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa act ata ggt aca gta tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggc tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aac ccg tac aat act cca gtc ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gat agt act aaa tgg aga aaa tta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aac aag aaa act caa gac ttc tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt cct tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag acg cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gcc ata nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn      816
Ala Ile Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
            260                 265                 270

nnn nnn nnn nnn nnn nnn nnn tat caa tac atg gat gat ttg tat gta      864
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gag aaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Lys Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gat aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gta ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc                                                                1059
Val
```

<210> 42
<211> 1053
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1053)
<223> Portion of HIV Reverse Transcriptase

<400> 42

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata arg ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Xaa Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atm aaa gta aga cag tat gat cag ata cyc ata     192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Xaa Ile
    50                  55                  60

gaa atc tgt gga yat aaa gct ata ggt acr gta tta gta gga ccc acg     240
Glu Ile Cys Gly Xaa Lys Ala Ile Gly Xaa Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac rta att gga aga aat ctg wtg act cag att ggt tgc act     288
Pro Val Asn Xaa Ile Gly Arg Asn Leu Xaa Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa ttr gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Xaa Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtc caa tta gga     576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aag aag aaa aaa tca gta aca gta ctg     624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg     672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt gta aac aat gag aca cca kgg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Xaa
225                 230                 235                 240

att aga tay cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata tty caa tgt agc atg aca aaa atc tta gag cct ttt aga aag      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac cta gtt att tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg ara tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Xaa Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg gca gtg caa cct ata gtg ctg cca gaa aaa gac agc tgg          1053
Lys Trp Ala Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp
            340                 345                 350


<210> 43
<211> 1082
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1082)
<223> Portion of HIV Reverse Transcriptase

<400> 43
cct caa atc act ctt tgg caa cga ccc ctt gtc aca rta aag rta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Xaa Lys Xaa Gly
 1               5                  10                  15

ggg caa cta aag gaa gct yta ttr gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Xaa Xaa Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
     50                  55                  60

gaa aty tgt ggg cat aaa gct ata ggt aca gta tta gta ggg cct aca      240
Glu Xaa Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc ccc aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aaa gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata ccg cat ccc gca ggg tta aaa aag aaa aag tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tat ast gca ttt acc ata ccg agt ata aac aat gag aca cca ggg      720
Lys Tyr Xaa Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt ccg cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa atc tta gaa cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285
```

```
gga tct gac ttg gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg ttg aaa tgg ggr ttt acc aca cca gac aag aaa cat      960
Arg Gln His Leu Leu Lys Trp Xaa Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa ccg ata gag ctg cca gaa aaa gaa agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Glu Leu Pro Glu Lys Glu Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gg                                 1082
Val Asn Asp Ile Gln Lys Leu Val
        355                 360


<210> 44
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 44
cct cag atc act ctt tgg caa cga ccc atc gtc aca gta aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct yta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aaa tgg aaa cca aaa ata ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt gcc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ala Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tka gga cat aaa gtt ata ggt aca gtc tta gta gga cct aca      240
Glu Ile Xaa Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aag att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aac agy act wga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Xaa Thr Xaa Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa ttr gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly
            180                 185                 190

ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg      624
Ile Pro His Pro Ser Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aga atc cta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg ctt gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val
        275                 280                 285

gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag ccc ata gtg ctg cca gaa aaa gay agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aag tta gtg gga aaa ttg aat tgg gca agc cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 45
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 45

```
cct cag atc act ctt tgg caa cga ccc rtc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gac gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aaa tgg aaa cca aaa atg ata gtg     144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Val
        35                  40                  45

gga att gga gga ttt gtc aaa gta aaa cag tat gag caa ata cct gta     192
Gly Ile Gly Gly Phe Val Lys Val Lys Gln Tyr Glu Gln Ile Pro Val
    50                  55                  60

gaa atc tgt gga cat aaa gct gta ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gar     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu
        115                 120                 125
```

```
aaa ata maa gca ttg gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Xaa Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gtg ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aag aac agt gat aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Asp Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag agg act caa gac ttc tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aag aaa tca gta aca rta cta      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Xaa Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca rtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Xaa Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa gct agc atg aca aaa atc tta gag cct ttc aga aaa      816
Ala Ile Phe Gln Ala Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa cta gtt atc tat caa tac gtg gat gac ttg tat gta      864
Gln Asn Pro Glu Leu Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg tta aaa tgg gga tta ttc aca cca gac cag aaa cat      960
Arg Glu His Leu Leu Lys Trp Gly Leu Phe Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg act ata cag cct atg gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Ile Gln Pro Met Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac cta cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Leu Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tat cca ggg                                                        1116
Ile Tyr Pro Gly
    370
```

<210> 46
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 46

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aaa gta ggg         48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta         96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga agg tgg aaa cca aaa atg ata ggg        144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata tcc ata        192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Ser Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca        240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act        288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag        336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gac ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa        384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gag att tgt aca gaa atg gaa aag gaa gga        432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aac cca tac aat act cca gta ttt        480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aaa gac agt act aag tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gag gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggc gat gca tat ttc tca gtt ccc tta gat gaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

act aga tat cag tac aat gtg ctc cca cag gga tgg aaa gga tca cca      768
Thr Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac cta gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc acc cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtr cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Xaa Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 47
<211> 1116

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 47

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg tgt ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Asp Met Cys Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga caa tat gat cag gta gcc atg       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Ala Met
    50                  55                  60

gaa atc tgt gga cat aag gct ata ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agc cct att gaa act gta ccm gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Xaa Val Lys Leu Lys
            100                 105                 110

cca ggr atg gat ggt cca agg gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Xaa Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata ara gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga       432
Lys Ile Xaa Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttt       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac tty tgg gaa gtt caa tta ggr       576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Xaa
            180                 185                 190
```

```
ata ccg cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctt      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg gga gat gca tat ttt tca gtt ccc tta gat aaa gat ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa ctr      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Xaa
    290                 295                 300

aga caa cat ctg ttg aag tgg ggg ytt acc aca cca gac aag aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Xaa Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccy cca ttc ctt tgg atg ggk tat gaa ctc cat cct gat     1008
Gln Lys Glu Xaa Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aar ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 48
<211> 1115
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 48

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

ata gaa gac ata gaa ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Ile Glu Asp Ile Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gag cag gta ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Gln Val Pro Ile
    50                  55                  60

gaa ctc tgt ggg cgt aaa act ata ggt aca gta tta gta gga cct aca     240
Glu Leu Cys Gly Arg Lys Thr Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aac ctg atg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aac act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcy ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc     528
Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga     576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aag aaa tca gta aca gta ttg     624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccg tta gat aaa gac ttc agg     672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctk cca cag gga tgg aag gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Xaa Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc ttg gag ccc ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac cta gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

ggc tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga ttt acc aca cca gat aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tcc car ga                                                     1115
Ile Ser Gln
    370
```

<210> 49
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

```
<400> 49
cct cag atc act ctt tgg caa cga ccc ctc gtc rca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Xaa Ile Lys Ile Gly
 1               5                  10                  15
```

```
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
               20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aag atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
          35                  40                  45

gga att gga ggt ttc atc aaa gta aga cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
      50                  55                  60

gaa atc tgt ggc cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat cta ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aag tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
           100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
      115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
  130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
           180                 185                 190

ata cca cat cct gca ggg tta aaa aag aam aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu
      195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
  210                 215                 220

aag tat acc gca ttt cca tcc cta gtt ata aac aat gag aca cca gga       720
Lys Tyr Thr Ala Phe Pro Ser Leu Val Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca       768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttt caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa gta gag gag ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Val Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agc cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 50
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 50

```
cct cag atc act ctt tgg caa cga ccc ttc gtc aac ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Asn Ile Lys Ile Gly
 1               5                  10                  15

gga caa ctg aag gaa gct cta ttg gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ttg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttk gtc aaa gta aga cag tat gat cag ata cct gta      192
Gly Ile Gly Gly Xaa Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
     50                  55                  60

gaa att tgt gga cat aaa gyt ata ggt aca gtc tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Xaa Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggc tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc ccg aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aag att ggg cct gaa aat cca tac aat act cca ata ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta mam aag aac aaa tca gta aca gtg cta      624
Ile Pro His Pro Ala Gly Leu Xaa Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta tat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Tyr Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tay aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc cag agt agc atg aca aga atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtc atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285
```

```
gca tct gac tta gaa ata gag aaa cat aga aca aaa ata gag gaa ctg      912
Ala Ser Asp Leu Glu Ile Glu Lys His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt tac aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gga ggg                                                   1116
Ile Tyr Gly Gly
    370
```

<210> 51
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 51

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ata ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt aty aaa gta aga cag tat gat cag ata cct ata      192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa act ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80
```

```
cct gcc aac ata att gga aga gat ctg ttg act cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa ttg aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gat ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cct gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gcg ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat gaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Glu Glu Phe Arg
    210                 215                 220

aag tat act gca ttc acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

gtt aga tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag ccc ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tat gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg agg tgg ggg ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320
```

```
cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg ggg aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 52
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 52

```
cct caa atc act ctt tgg caa cga ccc ctt gtc aca ata aag rta ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Xaa Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atr ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Xaa Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ycc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Xaa Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt tca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Ser Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata aty gga aga aat ctg atg act cag att ggt tgc act     288
Pro Val Asn Ile Xaa Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa ack gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Xaa Val Pro Val Lys Leu Lys
            100                 105                 110
```

```
cca gga atg gat ggc cca aaa gtt aag caa tgg cca ttg aca gra gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Xaa Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gag atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aga att ggg ccc gaa aat cca tac aat act cca ata ttt      480
Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aag aat agt act aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aac aaa tca gtg aca gta ytg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Xaa
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt atr aac aat gag aaa cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Lys Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca car gga tgg aaa ggg tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa aty tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Xaa Leu Glu Pro Phe Arg Lys
            260                 265                 270

car aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga aca aaa ata gag gaa ttg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tta agg tgg gga ttt ttc aca cca gaa caa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Phe Thr Pro Glu Gln Lys His
305                 310                 315                 320

cag aaa gaa ccg cca ttc ctt tgg atg ggt tat gaa cta cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg acg gta cag cct ata aag ctg cca gaa aaa gat agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350
```

```
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tay gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 53
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 53

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aaa gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gtg aga cag tat gat cag rta ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga tct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Ser Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggg tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gag atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140
```

```
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

atc cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc cgg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt agg aat      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Asn
            260                 265                 270

aaa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Lys Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac cta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg ttg aag tgg ggg ttt act aca cca gac aaa aaa cat      960
Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtc cag cct ata gag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gga                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 54
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 54

```
cct cag atc act ctt tgg caa cga ccc aty gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct yta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg gat ttg cca gga aga tgg aaa cca aaa atg ata gtg     144
Leu Glu Asp Met Asp Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Val
        35                  40                  45

gga att gga ggt ttt gtc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa att ata ggt aca gta tta ata gga aat aca     240
Glu Ile Cys Gly His Lys Ile Ile Gly Thr Val Leu Ile Gly Asn Thr
65                  70                  75                  80

cct gcc aac gta att gga aga aat ctg ttg act cag ctt ggt tgc act     288
Pro Ala Asn Val Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gat ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag gac agt act aaa tgg aga aaa gta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Val Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aac aag aga act caa gac ttc tgg gag gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cac ccc gca ggg ata aaa aag aat aaa tca gta act gta cta      624
Ile Pro His Pro Ala Gly Ile Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gta ggt gat gca tat ttc tca gtt ccc tta gat gaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aaa tat act gca ttc acc ata cct agt att aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctc cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cac aga ata aaa ata rag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ile Lys Ile Xaa Glu Leu
    290                 295                 300

aga gaa cat cta tgg aag tgg gga ttt tac aca cca gac aaa aag cat      960
Arg Glu His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata acg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Thr Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg ggg aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 55
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 55

```
cct caa atc act ctt tgg caa cga ccc ctc gtc gca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gtc        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aag cag tat gat cag gta ctt gta       192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Val Leu Val
    50                  55                  60

gaa att tgt gga cat ara gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Xaa Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgt act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca ggt atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt acc aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa acg caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205
```

```
gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aag gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctg cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga rca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Xaa Lys Ile Glu Glu Leu
    290                 295                 300

agg cag cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata ktg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tam ccc ngg                                                    1116
Ile Xaa Pro Xaa
    370
```

<210> 56
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 56

```
cct caa atc act ctt tgg caa cga ccc att gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata acc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Thr Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aaa gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gat agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gta caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttc acc ata cct agt ata aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agc agc atg aca aaa att tta gaa cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta raa ata gag cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Xaa Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa cag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Gln Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 57
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 57

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag tta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Leu Gly
 1               5                  10                  15

ggg caa cta atg gaa gtt cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Met Glu Val Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
rta gaa gaa ata agt tta cca gga aga tgg aaa cca aaa atg ata ggg      144
Xaa Glu Glu Ile Ser Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
          35                  40                  45

gga att gga ggt ttt gtc aaa gta aaa cag tat gat cag gta ccc tta      192
Gly Ile Gly Gly Phe Val Lys Val Lys Gln Tyr Asp Gln Val Pro Leu
      50                  55                  60

gaa att tgt gga aaa aag gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ttt ttg gct cag att ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Phe Leu Ala Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttc ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aga tgg aga aaa tta gta gat ttt      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag agg acs caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Xaa Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aar aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat cca gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Pro Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cca ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gac ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt tac aca cca gac caa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Tyr Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata acg ctg cca gac aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Thr Leu Pro Asp Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370


<210> 58
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 58
cct caa atc act ctt tgg caa cga ccc cta gtt aca ata aaa ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg act ttg cca gga aaa tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Thr Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga car tat gat cag ata ctc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60
```

```
gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag atc ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gag act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aar caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gag aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cca gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca ata atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Ile Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300
```

```
aga cag cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cca gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata aag ctg cca gac aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Asp Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gga                                                   1116
Ile Tyr Ala Gly
    370


<210> 59
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 59
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa
 1               5                  10                  15

ggg caa cta aaa gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat ttg cca ggg aaa tgg aaa cca maa atg ata ggg      144
Leu Glu Glu Ile Asn Leu Pro Gly Lys Trp Lys Pro Xaa Met Ile Gly
        35                  40                  45

gga att gga ggt ttt att aaa gta aga cag tat gat caa ata gcc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Ala Ile
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta rta gaa atc tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Xaa Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcm ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtc caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gac caa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Gln Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca agg atc tta gar cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtc aty tat cag tac atg gat gat tta tat gta      864
Gln Asn Pro Glu Ile Val Xaa Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa gta gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Val Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agr tgg ggg ttt tmc acg cca gac aaa aag cat      960
Arg Gln His Leu Leu Xaa Trp Gly Phe Xaa Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag act ata gaa ctg cca gaa aaa gat agc tgg act      1056
Lys Trp Thr Val Gln Thr Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

ata tac cca ggg                                                     1116
Ile Tyr Pro Gly
    370
```

<210> 60
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 60

```
cct caa atc act ctt tgg cag cga ccc cty gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aaa gaa gct cta tta gay aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca ggr aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Xaa Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata cct rta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Xaa
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg act cag ctt ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gag      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125
```

This page is not part of

the pamphlet!

# WO 01-35316

# 4/5

Date: 17 may 2001

Destination: Agent

Address:

```
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt aat aga tgg aga aaa tta gtg gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Asn Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aar aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta raa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Xaa Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt acc aat aat gag aca ccm ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Xaa Gly
225                 230                 235                 240

gtt aga tat cag tat aat gta ctt ccc cag gga tgg aaa gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca tat tty caa tgt agy atg aca aaa atc tta aag cct ttc agg aaa      816
Ala Tyr Phe Gln Cys Xaa Met Thr Lys Ile Leu Lys Pro Phe Arg Lys
            260                 265                 270

caa aat cca cac ata gtt att ttt caa tat gtg gat gac ttg tat gta      864
Gln Asn Pro His Ile Val Ile Phe Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gac tta gaa ata gag cag cat aga aca aaa ata gag gaa ctg      912
Ala Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ttg ttg agg tgg gga ctc acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

caa aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag ccc ata acg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Thr Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tat gca ggg                                                        1116
Ile Tyr Ala Gly
    370


<210> 61
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 61
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aaa gat agg          48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Asp Arg
 1               5                  10                  15

ggg gca agt aaa gaa gct cta tta gat aca gga gca gat gat aca gta          96
Gly Ala Ser Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat ttg cca ggg rag tgg aaa cca aaa atg ata ggg         144
Leu Glu Glu Ile Asn Leu Pro Gly Xaa Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tmt gat cag ata ccc gta         192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Xaa Asp Gln Ile Pro Val
    50                  55                  60

gaa att tgt gga cat aag gct ata ggt aca gta tta gta gga cct aca         240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag mtt ggt tgc act         288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Xaa Gly Cys Thr
                85                  90                  95

tta aat ttt ccc atc agt cct att gaa act gta cca gta aaa tta aag         336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca tta aca gag gaa         384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg         432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt         480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt cag tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa agc ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Ser Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata ccc agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

rca aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Xaa Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa atg gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gag ata gag caa cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 62
<211> 1116

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 62

```
cct caa atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga caa tat gat cag ata gcc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Ala Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg act cag att ggc tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gat act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Asp Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aat agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc gca ggg cta aaa aag aay aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtc ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

rtt aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Xaa Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

tca ata ttc caa tgt agc atg acg aaa atc tta gag cct ttt aga aaa      816
Ser Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtt atc trt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Xaa Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gac tta gaa ata gag cag cat aga ata aaa ata gag gaa cta      912
Ala Ser Asp Leu Glu Ile Glu Gln His Arg Ile Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga ttt tac aca cca gac aaa aaa yat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys Xaa
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gar ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag ttr gtg gga aaa ctg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Xaa Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 63
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse-Transcriptase

<400> 63

```
cct caa atc act ctt tgg caa cga ccc gtt gtt aca gta agg ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Val Arg Ile Gly
 1               5                  10                  15

gga cag cta acg gaa gct yta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Thr Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg act ttg cca gga aaa tgg aaa cca aaa ata ata ggg       144
Leu Glu Glu Met Thr Leu Pro Gly Lys Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

ggr att gga ggt ttt atc aaa gta aga cag tat gat cac gta ctt gta       192
Xaa Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp His Val Leu Val
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg atg act cag ctt ggg ttc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Phe Thr
                85                  90                  95

tta aat ttt cca att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca ggg atg gat ggc cca aaa gtt aaa caa tgg cca ttg mca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Xaa Glu Glu
        115                 120                 125

aaa ata aaa gca cta aca gaa att tgt aca gaa ttg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aga ata ggg cct gaa aat cca tac aat act cca ata ttt       480
Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aag aaa aac ggt ayt agg tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Gly Xaa Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gag cta aat aag aga act caa gac ttc tgg gaa gtt caa cta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca gga cta aaa aag aac aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta cat gaa gac ttt aga       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu His Glu Asp Phe Arg
    210                 215                 220
```

```
aag tat acc gca ttc acc ata cct agt aca aac aat gaa aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca ccg      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg acc aaa atc tta gaa cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa atg gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Met Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga ata aaa ata gag gaa tta      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ile Lys Ile Glu Glu Leu
    290                 295                 300

agg gaa cac cta ttg aag tgg gga ttt ttc acc cca gac gaa aag cat      960
Arg Glu His Leu Leu Lys Trp Gly Phe Phe Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctt cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata aaa ctg cca gaa aaa gaa agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Glu Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa tta aat tgg gca agc cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca gga                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 64
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 64

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15
```

```
ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aaa tgg aaa cca aaa atr ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Xaa Ile Gly
        35                  40                  45

gga att gga ggy ttt rtc aaa gta aga cag tat gat cag ata syc ata       192
Gly Ile Gly Xaa Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Xaa Ile
    50                  55                  60

gaa atc tgc gga cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gyc aac ata att gga aga aat ctg ttg act cag ctt ggg tgc act       288
Pro Xaa Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta caa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Gln Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca tta aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aag ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aag gac agt gct aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Ala Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta ggg       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cck cat ccc gca ggg ttr aaa aag aaa aaa tca gta aca gta cta       624
Ile Xaa His Pro Ala Gly Xaa Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gta ggt gat gca tat ttt tca gtt ccc tta gat caa aac ttc aga       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Gln Asn Phe Arg
    210                 215                 220

aag tat act gca ttc acc ata cct agt ata aac aat gag ayg cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Xaa Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca       768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gar ata rtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Xaa Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac ttr gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Xaa Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ytg ttg aag tgg gga ttt acc aca cca gac aag aag cat      960
Arg Gln His Xaa Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata atg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Met Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gga                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 65
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 65

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atc aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg      144
Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttt gtc aaa gta aga gag tat gat cag gta ccc ata       192
Gly Ile Gly Gly Phe Val Lys Val Arg Glu Tyr Asp Gln Val Pro Ile
     50                  55                  60

gac atc tgt gga cat aaa gtt ata ggt aca gtg tta gta gga cct aca       240
Asp Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act       288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gar atc tgt aca gaa ttg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aay cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctr       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Xaa
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

rtt aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca       768
Xaa Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aaa       816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata att atc tat caa tac gtg gat gat ttg tat gta       864
Gln Asn Pro Asp Ile Ile Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285
```

```
gga tct gat ttg gaa ata gag cag cat aga aca aaa ata gag gaa cta      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Glu His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aag gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata aag ytg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Xaa Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                    1116
Ile Tyr Pro Gly
    370


<210> 66
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 66
cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gak rca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Xaa Xaa Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agr car tat gac cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cag aaa gct ata ggt aca gta tta gta gga cct acm      240
Glu Ile Cys Gly Gln Lys Ala Ile Gly Thr Val Leu Val Gly Pro Xaa
65                  70                  75                  80
```

```
cct gtc aac ata att gga aga aat ctg ttg act caa att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gca gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Ala Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt aat ara tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Asn Xaa Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctc aat aag aga act caa gac ttc tgg gaa gtt caa tta ggc      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aam aaa tca gta aca rta ctr      624
Ile Pro His Pro Ala Gly Leu Lys Lys Xaa Lys Ser Val Thr Xaa Xaa
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aar tat act gca ttt acc ata cct agt aca wac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Xaa Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag krc aat gtg yyt cca cag gga tgg aaa gga tcm cca      768
Ile Arg Tyr Gln Xaa Asn Val Xaa Pro Gln Gly Trp Lys Gly Xaa Pro
                245                 250                 255

gca ata ttc mam agt agc ayg aca aaa att tta gag cct ttt aga aaa      816
Ala Ile Phe Xaa Ser Ser Xaa Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tgt caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Cys Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ttg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

agg caa cat ttg ttg agg tgg ggr ttt acc aca cca gac ara aaa cat      960
Arg Gln His Leu Leu Arg Trp Xaa Phe Thr Thr Pro Asp Xaa Lys His
305                 310                 315                 320
```

```
cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata aaa ctg cca gaa aaa gay agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 67
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase.

<400> 67

```
cct caa atc act ctt tgg caa cga cca ata gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

cta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt aty aaa gta aga cag tat gat cag ata tcc ata     192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Ser Ile
    50                  55                  60

gaa atc tgt ggg cat aaa gtt aca ggt aca gtg tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Val Thr Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110
```

```
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttg gta gaa att tgt gca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Ala Glu Met Glu Lys Glu Gly
    130                 135                 140

caa att tca aaa att gag cct gaa aat cca tac aat aat cca gta ttt      480
Gln Ile Ser Lys Ile Glu Pro Glu Asn Pro Tyr Asn Asn Pro Val Phe
145                 150                 155                 160

gtc ata aag aaa aaa gac ggt act aac tgg aga aaa tta ata gat ytc      528
Val Ile Lys Lys Lys Asp Gly Thr Asn Trp Arg Lys Leu Ile Asp Xaa
                165                 170                 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aat aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca ttt tat tca gtt ccc tta gat gag aac ttc agg      672
Asp Val Gly Asp Ala Phe Tyr Ser Val Pro Leu Asp Glu Asn Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca atg gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Met Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

aac aat cca gac ata gtc atc tat caa tac atg gat gat ttg tat gta      864
Asn Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gat tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Ala Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat cta ttr aag tgg gga ttt acc aca cca gac aar aar yat      960
Arg Glu His Leu Xaa Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys Xaa
305                 310                 315                 320

cag aaa gaa cct cca ytc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Xaa Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350
```

```
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg att                                              1119
Ile Tyr Pro Gly Ile
    370
```

<210> 68
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 68

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

gga caa cta aaa gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca ggg aaa tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga atc gga gga ttt atc aaa gta aga cag tat gag cag ata cac ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Glu Gln Ile His Ile
    50                  55                  60

gaa atc tgt ggg cat aaa gct ata ggt aca gtr tta ata gga ccc aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Xaa Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gag      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140
```

```
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gtt ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg ttg aag aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa aac ttt agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asn Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aat aat gaa aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa gct agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ala Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac atg rtt att tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Xaa Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

ggc tct gac tta gaa ata gga cag cat aga aca aaa ata gaa gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctc tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gcg agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg att                                               1119
Ile Tyr Pro Gly Ile
    370
```

<210> 69
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 69

```
cct cag atc act ctt tgg caa cga ccc cty gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa yta aag gaa gct mta tta gay aca gga gca gat gat aca gtg      96
Gly Gln Xaa Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ata ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga gag tat gag cag ata caa gta     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Glu Tyr Glu Gln Ile Gln Val
    50                  55                  60

gaa atc tgt gga cat aag gct ata rgt aca gta tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Xaa Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat cta atg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gag act gta ccg gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aga gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat acy ccr gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Xaa Xaa Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata ccg cat ccc gca ggg tta aag aag aaa aaa tca gta aca gta ctr      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Xaa
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gaa cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat car tac atg gat gac ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa cta      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tkg agg tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Xaa Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cac cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctr cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Xaa Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gcg agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat tca ggg att                                               1119
Ile Tyr Ser Gly Ile
    370
```

<210> 70
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 70

```
cct caa atc act ctt tgg caa cga ccc cty gtc kca ata aag gta ggr        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Xaa Ile Lys Val Xaa
 1               5                  10                  15

ggg caa mta aag gaa gct yta tta gat aca gga gca gat gat aca gta        96
Gly Gln Xaa Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gat cag gta arc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Val Xaa Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aay ctg ttg aca cag att ggt tgy act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca ara gty aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Xaa Xaa Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aar gca tta atg gaa att tgt gca gay atg gaa aag gaa ggr       432
Lys Ile Lys Ala Leu Met Glu Ile Cys Ala Asp Met Glu Lys Glu Xaa
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcy ata aag aaa aaa gac agc act aaa tgg aga aaa tta gta gat ttc       528
Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttt tgg gaa gtc caa tta gga       576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccy gca ggg tta aaa aag aac aaa tca gta aca gta ttg       624
Ile Pro His Xaa Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205
```

```
gat gtg ggt gat gca tat ttt tca gtt ccy tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Xaa Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tay act gca ttt acm ata cct agt ata aat aat gca aca cca ggg      720
Lys Tyr Thr Ala Phe Xaa Ile Pro Ser Ile Asn Asn Ala Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga rar      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Xaa
            260                 265                 270

cag aat cca gac ata gtt atc tat caa tac atg gat gay ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa mta ggg cag cat aga rca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Xaa Gly Gln His Arg Xaa Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tta agg tgg ggg ttt acc acw cca gac aag aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Xaa Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta car ccc ata gtg ttg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tay gsa ggg att                                               1119
Ile Tyr Xaa Gly Ile
    370
```

<210> 71
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 71

```
cct caa atc act ctt tgg caa cga ccc atc gtc tca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Ser Ile Lys Ile Gly
 1               5                  10                  15

ggg gca aat aaa gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Ala Asn Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aag cca aaa atg ata gtg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Val
        35                  40                  45

gga att gga ggt ttt agc aaa gta aga caa tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ser Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgc gga cgt aaa gtt gta ggt tca gta tta ata gga cct aca       240
Glu Ile Cys Gly Arg Lys Val Val Gly Ser Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag ctt ggc tgt act       288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gag       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt aca gaa ttg gaa gaa gma gga       432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Leu Glu Glu Xaa Gly
    130                 135                 140

aaa att aca aaa att ggg cct gaa aat ccg tac aat act cca ata ttt       480
Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aar aac agt act aaa tgg aga aaa tta gta gac ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat aaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aar tat act gca ttt acc ata cct agt acg aat aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat ccc gac ata gtt atc tat caa tac gtg gat gat ttg ctt gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gag gag cta      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata caa aag tta gtg gga aaa tta aat tgg gcw agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Xaa Ser Gln
        355                 360                 365

att tat cca ggg att                                                1119
Ile Tyr Pro Gly Ile
    370
```

<210> 72
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 72

```
cct cag atc act ctt tgg caa cga ccc cty gtc aca ata aag atc ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa tta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
ata gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Ile Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
         35                  40                  45

gga att gga ggt ttt rtc aaa gta aga caa tat gat cag gta ccc ata      192
Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
     50                  55                  60

gaa att tgc gga cat aaa gct ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
 65                  70                  75                  80

cct gyc aac ata att gga aga aac ctg ttg act caa ctt ggc tgc act      288
Pro Xaa Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt cca att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aaa gga agg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Gly Arg
    130                 135                 140

aaa aat tac aaa att ggg cct gaa aac cca tac aat act cca gta ttt      480
Lys Asn Tyr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat aag gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agc ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctc cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gcm ata ttc caa tgt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Xaa Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

ggg tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cga cat ctg ttg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Arg His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gag ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta caa cct ata gtg cta cca gag aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aag tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

ata tac gca ggg att                                                1119
Ile Tyr Ala Gly Ile
    370
```

<210> 73
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 73

```
cct caa atc act ctt tgg caa cga ccc ttc gtc aca gta aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat aat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asn Thr Val
            20                  25                  30

tta gaa gaa atg aat tta ccg gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag rta ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile
    50                  55                  60
```

```
gaa atc tgt gga cac aaa gct ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga gat ctg ttg act cag ctt ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gat act gta cca gta aaa tta aaa      336
Leu Asn Phe Pro Ile Ser Pro Ile Asp Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat acc cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aag tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gcg ggg tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt ccc cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtt atc tac caa tac gtg gat gac ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gat gag ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Asp Glu Leu
    290                 295                 300
```

```
agg caa cat ctg ttg aag tgg gga ttt tac aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cca cca ttc ctt tgg atg ggk tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aar gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg att                                               1119
Ile Tyr Pro Gly Ile
    370


<210> 74
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 74
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag gtc ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gag gaa cta aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Leu Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa ata tgt gga cat aaa gct att ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aac ttg ttg act cag ctt ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta aca gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggt tta aaa aag aaa aaa tca gta aca gtc ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tac cag tac aat gtg ctt ccc cag ggg tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tac caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gag ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gat agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                     1116
Ile Tyr Ala Gly
    370
```

<210> 75
<211> 819
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(819)
<223> Portion of HIV Reverse Transcriptase

<400> 75

```
ccc att agt cct att gam act gta cca gta aaa tta aag cca gga atg       48
Pro Ile Ser Pro Ile Xaa Thr Val Pro Val Lys Leu Lys Pro Gly Met
 1               5                  10                  15

gat ggc cca aaa gtt aaa caa tgg cca tta aca gag gaa aaa ata aaa       96
Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys
            20                  25                  30

gca ttg gta gaa att tgt aca gaa atg gaa aag gaa gga aaa att tca      144
Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser
        35                  40                  45

aaa att ggg cct gaa aat cca tac aat act cca gta ttt gcc ata aag      192
Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys
    50                  55                  60

aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc aga gaa ctt      240
Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu
65                  70                  75                  80

aat aar aga act caa gat ttc tgg gaa gtt caa tta gga ata cca cat      288
Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His
                85                  90                  95

ccc tca ggg tta aaa aag aay aaa tca gta aca gta ttg gat gtg ggt      336
Pro Ser Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu Asp Val Gly
            100                 105                 110

gat gca tat ttt tca gtt ccy tta gat aaa gac ttc agg aag tat act      384
Asp Ala Tyr Phe Ser Val Xaa Leu Asp Lys Asp Phe Arg Lys Tyr Thr
        115                 120                 125

gca ttt acc ata cct agt ata aac aat gag aca cca ggg att agr tat      432
Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Xaa Tyr
    130                 135                 140
```

```
cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca gca ata ttc      480
Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe
145                 150                 155                 160

caa agt agc atg aca aaa atc tta gag cct ttt aga aaa cat aat cca      528
Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys His Asn Pro
                165                 170                 175

gac ata gtt atc tat caa tac gtg gat gat ttg tat gta gga tct gac      576
Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val Gly Ser Asp
            180                 185                 190

tta gaa ata gag gag cat aga aca aaa ata gag gaa ctg agr vrg cat      624
Leu Glu Ile Glu Glu His Arg Thr Lys Ile Glu Glu Leu Xaa Xaa His
        195                 200                 205

ctg tta aag tgg gga ttt acy aca cca gac aaa aag cat cag aaa gaa      672
Leu Leu Lys Trp Gly Phe Xaa Thr Pro Asp Lys Lys His Gln Lys Glu
    210                 215                 220

cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat aaa tgg aca      720
Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr
225                 230                 235                 240

gta cag cct ata aag ctg cca gaa aaa gac agc tgg act gtc aat gac      768
Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp
                245                 250                 255

ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag att tat gca      816
Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala
            260                 265                 270

ggg                                                                 819
Gly
```

<210> 76
<211> 819
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(819)
<223> Portion of HIV Reverse Transcriptase

<400> 76

```
ccc att agt cct att gaa act gta cca gta aaa tta aag cca gga atg       48
Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met
 1               5                  10                  15

gat ggc cca aaa gty aaa caa tgg cca tta aca gaa gaa aaa ata aga       96
Asp Gly Pro Lys Xaa Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Arg
            20                  25                  30

gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga aaa att tca      144
Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser
        35                  40                  45
```

```
aaa att ggg cct gaa aat cca tac aat act cca gtg ttt gct ata aag      192
Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys
     50                  55                  60

aaa aaa gac agt act aar tgg aga aaa ttg gta gat ttc aga gaa ctt      240
Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu
 65                  70                  75                  80

aat aag aga act caa gac ttc tgg gaa gtt caa tta gga ata cca cat      288
Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His
                85                  90                  95

ccc tca ggg tta aaa aag aaa aaa tca gta aca gta ctg gat gtg ggt      336
Pro Ser Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly
            100                 105                 110

gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg aag tat act      384
Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg Lys Tyr Thr
        115                 120                 125

gca ttt act atn cct agt ata aac aat gag aca cca ggg att agg tat      432
Ala Phe Thr Xaa Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr
    130                 135                 140

cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca gca ata ttc      480
Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe
145                 150                 155                 160

caa agt agc atg aca aaa atc tta gag cct ttt aga aaa caa aat cca      528
Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro
                165                 170                 175

gac ata gtt atc tat caa tac gtg gat gat ttg tat gta gga tct gac      576
Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val Gly Ser Asp
            180                 185                 190

cta gaa ata gga cag cat aga aca aaa ata gag gaa ctg aga cag cat      624
Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His
        195                 200                 205

ctg ttg agg tgg gga ttt acc aca cca gac aag aaa cat cag aaa gaa      672
Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu
    210                 215                 220

cct ccc ttt ctt tgg atg ggc tat gaa ctc cat cct gat aaa tgg aca      720
Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr
225                 230                 235                 240

gta cag cct ata gag ctg cca gac aag gat agc tgg act gtc aat gac      768
Val Gln Pro Ile Glu Leu Pro Asp Lys Asp Ser Trp Thr Val Asn Asp
                245                 250                 255

ata cag aag tta gtg gga aaa tta aat tgg gca agt cag ata tat gca      816
Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala
            260                 265                 270

ggg                                                                  819
Gly
```

<210> 77
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 77

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg aat ttg cca ggg aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Asp Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata cct ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgc gga cat aaa gct gta ggt aaa gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Val Gly Lys Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act caa ctt ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aag caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gag aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa aac agt act aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aat aag aga acg caa gac ttc tgg gaa gtt caa nnn nnn      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Xaa
            180                 185                 190

nnn nnn nnn nnn nnn ggg twa aaa aag aaa aaa tca gta aca gta ctg      624
Xaa Xaa Xaa Xaa Xaa Gly Xaa Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gta ggt gat gca tat ttc tca gtt cct cta gat aaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttc acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctg cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gtg      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ttg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agc cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

ata tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 78
<211> 1122
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1122)
<223> Portion of HIV Reverse Transcriptase

<400> 78

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg gat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Met Asp Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttg gta gaa ata tgt aca gaa atg gaa aag gaa ggg     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat acr cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Xaa Pro Val Phe
145                 150                 155                 160

gcc ata arg aaa aaa gaa agc tct agc tct aaa tgg aga aaa tta gta     528
Ala Ile Xaa Lys Lys Glu Ser Ser Ser Ser Lys Trp Arg Lys Leu Val
                165                 170                 175

gat ttc aga gaa ctt aat aar aga act caa gac ttt ttk gaa gtt caa     576
Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Xaa Glu Val Gln
            180                 185                 190

tta gga ata cca cat ccc gca ggg tta aag aag aaa aaa tca gya aca     624
Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Xaa Thr
        195                 200                 205
```

```
rta ttg gat gtg ggt gat gca tat ttt tca gtt ccc tta gat raa gac      672
Xaa Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Xaa Asp
    210                 215                 220

ttc agg aag tat act gca ttt acc ata cct agt ata aac aat gag aca      720
Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr
225                 230                 235                 240

cca ggg att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga      768
Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly
                245                 250                 255

tca cca gct ata ttc caa agt agc atg aca aaa atc tta gag cct ttt      816
Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe
            260                 265                 270

aga aaa caa aat cca gay ata gtt atc tat caa tac atg gat gat ttg      864
Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu
        275                 280                 285

tat gta gga tct gay tta gaa ata gag cag cat aga ata aaa ata gag      912
Tyr Val Gly Ser Asp Leu Glu Ile Glu Gln His Arg Ile Lys Ile Glu
    290                 295                 300

gaa ctg aga caa yat ytg tgg arg tgg ggr ttt tac aca cca gac aaa      960
Glu Leu Arg Gln Xaa Xaa Trp Xaa Trp Xaa Phe Tyr Thr Pro Asp Lys
305                 310                 315                 320

aaa cat cag aaa gaa cct cca ttc cat tgg atg ggt tat gaa ctc cat     1008
Lys His Gln Lys Glu Pro Pro Phe His Trp Met Gly Tyr Glu Leu His
                325                 330                 335

cct gat aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc     1056
Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser
            340                 345                 350

tgg act gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca     1104
Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala
        355                 360                 365

agt cag att tat gca ggr                                            1122
Ser Gln Ile Tyr Ala Xaa
    370
```

<210> 79
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 79

```
cct cag atc act ctt tgg caa cga ccc ctc gtt aca ata aag gta ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gac aat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asn Thr Val
            20                  25                  30

ttc gaa gac ctg gat tta cca gga agg tgg aaa cca aaa atg ata ggg       144
Phe Glu Asp Leu Asp Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gag cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt ggg cgt aaa gct ata ggt aca gtg tta gta gga cct aca       240
Glu Ile Cys Gly Arg Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga gat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

cta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa att tgt gca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Ala Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt aat aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Asn Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aag tca ata aca gta tta       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Ile Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat gaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
att aga tat cag tac aat gtg ctg cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata gag cag cat aga aca aaa ata gat gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Asp Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ctt acc aca cca gac cag aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gac aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Asp Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg ggr aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Xaa Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 80
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 80

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gag gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg acw cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Xaa Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aar gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca rta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Xaa Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag agg act caa gat ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg ttg aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat gaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gac ata gtc atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat agg aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ttg ttg aag tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata gtg tta ccg gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg att                                                1119
Ile Tyr Pro Gly Ile
    370
```

<210> 81
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 81

```
cct caa atc act ctt tgg caa cga ccy ctt gtt rcc ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Xaa Leu Val Xaa Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta arg gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Xaa Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aaa cag tat gat caa ata ccy rta      192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Xaa Xaa
    50                  55                  60
```

```
gaa att tgt gga cat aga gct ata ggt aca gtw tta gta gga cct aca      240
Glu Ile Cys Gly His Arg Ala Ile Gly Thr Xaa Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga agr aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Xaa Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttg gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aga att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aar gat agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag agg act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt act ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cca ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtc atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa yta      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Xaa
    290                 295                 300
```

```
aga gaa cat ctg tta arg tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Glu His Leu Leu Xaa Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata cag ctg cca gaa aag gaa agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys Glu Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 82
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 82

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa tta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agt tta cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctt gta     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Val
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

ccc gtc aac ata att gga aga aat ctg ttg act cag att ggg tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aaa aag aaa gac agt act aaa tgg aga aag tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aay aaa aag act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gam ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Xaa Phe Arg
    210                 215                 220

aar tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tay cag tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggr aag cac aga aca aaa ata gag gag cta      912
Gly Ser Asp Leu Glu Ile Xaa Lys His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctk tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Xaa Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag cct ata aaa ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gty aat gac ata cag aag tta gtg gga aaa ttr aat tgg gcc agt cag     1104
Xaa Asn Asp Ile Gln Lys Leu Val Gly Lys Xaa Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 83
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 83

```
cct cag atc act ctt tgg caa cga cca ctc gtc gca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac atg agt tta cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Asp Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat caa gta ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125
```

```
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tat aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg agg aaa tta gta gat ttt      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cca gca ggg tta aaa aag aaa aag tca gta aca gtg ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata ccc agt ata aac aat gag aca ccc agg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Arg
225                 230                 235                 240

gtt aga tat caa tac aat gta ctt cca cag gga tgg aaa gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca tat ttc caa agt agc atg aca aaa atc tta gaa ccc ttc aga aaa      816
Ala Tyr Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aac cca gac ata gtt atc tat caa tac atg gat gac tta tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gag ata gga cag cat aga gca aaa ata gag gac cta      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Asp Leu
    290                 295                 300

aga gca cat ctg ttg aag tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Ala His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttt ctc tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gwg cta cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gta gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tat cca ggg                                                        1116
Ile Tyr Pro Gly
    370


<210> 84
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 84
cct caa atc act ctt tgg caa cga ccc att gtc aca ata aaa gta ggg         48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta atg gaa gct cta tta gat aca gga gca gat gat aca gta         96
Gly Gln Leu Met Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata aat ttg cca gga aga tgg aaa cca aaa ata ata ggg        144
Leu Glu Asp Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att ggt ggt ttt gtc aaa gtg aga cag tat gat cag gta ccc ata        192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca        240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct acc aac gta gtt gga aga aat ctg atg act cag att ggc tgc acy        288
Pro Thr Asn Val Val Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Xaa
                85                  90                  95

tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag        336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg acg gaa gaa        384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gat gga        432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tat aat act cca ata ttt        480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aag aac agt gat aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Asp Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aar aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aat aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat ata ggt gat gca tat ttt tca att ccc tta gat aaa gac ttt agg      672
Asp Ile Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttc acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

gtt aga tat cag tac aat gtg ctt cca cag gga tgg aag gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agc agc atg acc aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtt atc tgc caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Cys Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctr      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Xaa
    290                 295                 300

agg aat yat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa tat      960
Arg Asn Xaa Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys Tyr
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag ccc ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 85
<211> 1116

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 85

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aaa gta ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca ggg aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta agc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Ser Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga ccc acc       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gag aag gaa ggr       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Xaa
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aar aaa aaa gac agt act aaa tgg aga aag tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa ara act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Xaa Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc gca ggg tta aaa aag aam aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu
        195                 200                 205

gay gtg ggt gat gcr tat ttt tca gtt ccy tta gay aaa gay ttc agg      672
Asp Val Gly Asp Xaa Tyr Phe Ser Val Xaa Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac aca gca ttt acc ata cct agt gta aac aat gag rca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Xaa Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca car gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aar      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

maa aat cca gac ata gty atc tay caa tac atg gat gat ttr tat gta      864
Xaa Asn Pro Asp Ile Xaa Ile Tyr Gln Tyr Met Asp Asp Xaa Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gga cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg cag tgg ggg tta acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Gln Trp Gly Leu Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat ccg gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata wtg ctg cca gac aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Asp Lys Asp Ser Trp Thr
            340                 345                 350

gtm aat gac ata cag aar tta gta gga aaa ttg aat tgg gcg agt cag     1104
Xaa Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

atc tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 86
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 86

```
cct caa atc act ctt tgg caa cga ccc atc gtc aca gta aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg cac aca acg gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly His Thr Thr Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca ggg aga tgg aaa cca aaa atg ata gga      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gag cag gta ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Glu Gln Val Pro Ile
    50                  55                  60

gaa ttc tgt gga cat aaa act gta ggt aca gta tta ata gga cct aca      240
Glu Phe Cys Gly His Lys Thr Val Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg act cag att ggt tgt act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggg ccc aaa gtt aaa cca tgg cca ttg aca gaa aga      384
Pro Gly Met Asp Gly Pro Lys Val Lys Pro Trp Pro Leu Thr Glu Arg
        115                 120                 125

aaa aat aaa gca tta gta gaa att tgt tcc gaa atg gaa aaa gga agg      432
Lys Asn Lys Ala Leu Val Glu Ile Cys Ser Glu Met Glu Lys Gly Arg
    130                 135                 140

aaa att tca aaa att ggg cct gag aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt cag tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gta ggt gat gca tat ttt tca gtt ccc tta gat gaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Glu Phe Arg
    210                 215                 220
```

```
aag tat act gca ttc acc ata cct agt aca aac aat gaa aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa atc tta gag ccc ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tgt cag tac atg gat gac ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gat tta gaa ata ggg cag cat aga aca aaa gta gag gaa ctg      912
Ala Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Val Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg ggg ttt ttc aca cca gac gaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Phe Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gta ctg cca gac caa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Asp Gln Asp Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa tta aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 87
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 87

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata gag      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Glu
 1               5                  10                  15
```

```
ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg tca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Ser Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gag atc tgt gga cat aaa gct gta ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga agr aat ctg ttg act cag att ggt tgc acc       288
Pro Val Asn Ile Ile Gly Xaa Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat tty       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccy gca ggg ttg aar aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Xaa Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat gaa gay ttc aga       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

gtt aga tat car tac aat gtg ctt cca cag gga tgg aag gga tca cca       768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttc caa agc agc atg aca aaa atc tta gag cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gat ata gtt atc tat caa tac atg gat gac ttr tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Xaa Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg car cat aga aca aaa ata gag gaa ttg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga tta acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Leu Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 88
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 88

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta agg raa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Arg Xaa Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata gaa ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Asp Ile Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gga ggt ttt gtc aaa gta aga caa tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
     50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg atg act cag ctt ggt tgc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu
        115                 120                 125

aaa ata gaa gca tta atr gaa att tgt gma ttt ttg gaa aag gaa gga      432
Lys Ile Glu Ala Leu Xaa Glu Ile Cys Xaa Phe Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat ccg tac aac act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gga ggt act aaa tgg aga aaa ata gta gat ttc      528
Ala Ile Lys Lys Lys Gly Gly Thr Lys Trp Arg Lys Ile Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gcg ggg tta aaa aag aay aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat gaa gaa ctc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Glu Glu Leu Arg
    210                 215                 220

aag tat act gca ttt act ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tac caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aaa atc tta gag ccc ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc twt caw tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Xaa Xaa Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285
```

```
gga tct gac tta gaa ata ggg aag cat agg gaa aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Lys His Arg Glu Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aag tgg gga ttt tac aca cca gac gaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat ctt gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Leu Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370


<210> 89
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 89
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agt ttg cca ggg aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Ser Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga caa ttt gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Phe Asp Gln Ile Pro Ile
    50                  55                  60

gaa ata tgt gga cac aaa gct ata ggt aca gta tta ata gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80
```

```
cct gtc aac ata att gga agg aat ctg ttg act cag ctt ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc atc agt cct att gaa cct gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Pro Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aaa gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctg aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta acg gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa cat agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln His Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg ttg aag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Glu His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320
```

```
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat      1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                     1116
Ile Tyr Ala Gly
    370
```

<210> 90
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 90

```
cct cag atc act ctt tgg caa cga ccc aty gtc aca ata aaa gta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly
 1               5                  10                  15

gga cag cta aag gaa gct yta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aac ttg cca gga aaa tgg aaa cca aaa ata ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aga gta aga caa tat gat cag gta cct gta      192
Gly Ile Gly Gly Phe Val Arg Val Arg Gln Tyr Asp Gln Val Pro Val
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt tca gta tta gta gga cca aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Ser Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg atg act cag ctt ggt ttc act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Phe Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110
```

```
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gar att tgt aca gaa ytg gaa aaa gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Xaa Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt gat aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Asp Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gga ggg tta aaa aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Gly Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat car tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata tty caa agt agc atg aca aaa atc tta gag cct ttt agg aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

maa aat cca gac ata gtt atc att caa tac atg gat gat ttg tat gtr      864
Xaa Asn Pro Asp Ile Val Ile Ile Gln Tyr Met Asp Asp Leu Tyr Xaa
        275                 280                 285

gga tct gat tta gaa ata gar cag cay aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gat cat tta ttg agg tgg ggg ttt ttc aca cca gaa caa aaa cat      960
Arg Asp His Leu Leu Arg Trp Gly Phe Phe Thr Pro Glu Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc cat tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe His Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cat cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val His Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350
```

```
gtc aat gac ata cag aag tta gtg gga aaa ttr aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Xaa Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 91
<211> 1115
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 91

```
cct cag atc act ctt tgg caa cga ccc ctt gtc aca gta aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

ttg gaa gaa atg aat ttg cca ggg aga tgg aaa cca aaa ata ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata rgt cca gta tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Val Ile Xaa Pro Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ttg atg act cag att ggc tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc atc agt cct att raa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Xaa Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aag gtt aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140
```

```
aaa atc tca aaa att ggg cct gaa aac cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aac agt act aga tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gga ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Gly Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt cct cta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aat aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

gtt aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tcg cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt cag gct agc atg aca aaa atc tta gag ccg ttt aga aaa      816
Ala Ile Phe Gln Ala Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac cta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ttg ttg aaa tgg gga ttt atc aca cca gat gaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Ile Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aag tgg aca gta cag cct ata gta ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gg                                                     1115
Ile Tyr Ala
    370
```

<210> 92
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 92

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gac ata aac ttg cca gga aaa tgg aaa cca aaa atg ata ggg       144
Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gag cag gta ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Glu Gln Val Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa act ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg atg act cag att ggg tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt act aga tgg aga aaa gta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Val Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag acg cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa ata tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ctg gtt atc tgt caa tac atg gat gat tta tat gta      864
Gln Asn Pro Asp Leu Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac cta gaa ata ggg cag cat aga aca aaa ata gaa gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

agg caa cat ctg ttg aag tgg gga ttt acc aca cca gac gaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Glu Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag ccc ata gtg ctg cca gac aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Asp Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 93
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 93

```
cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata gga      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta ata gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat tta cca gga aga tgg aca cca aaa ata ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Thr Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt ttt gtc aga gta aga cag tat gaa cag ata ccc gta     192
Gly Ile Gly Gly Phe Val Arg Val Arg Gln Tyr Glu Gln Ile Pro Val
    50                  55                  60

gaa atc tgc ggg cat aaa gct gta ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggc tgt act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gat act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Asp Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca ara gtt aaa caa tgg cca ttg aca gaa gag     384
Pro Gly Met Asp Gly Pro Xaa Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gam gga     432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Xaa Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt     480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa gta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Val Val Asp Phe
                165                 170                 175

aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga     576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg ata maa aag aac aaa tca gta aca gta ytg     624
Ile Pro His Pro Ala Gly Ile Xaa Lys Asn Lys Ser Val Thr Val Xaa
        195                 200                 205
```

```
gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gag gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gta ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa aty tta gag cct ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Xaa Leu Glu Pro Phe Arg Lys
            260                 265                 270

aaa aat cca gac ata rtt atc tgc caa tac atg gat gat ttg tat gta      864
Lys Asn Pro Asp Ile Xaa Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gag cag cat aga aca aaa ata gat gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Asp Glu Leu
    290                 295                 300

aga gac cat ctg tgg aag tgg gga ttt tac aca cca gac aac aaa yat      960
Arg Asp His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Asn Lys Xaa
305                 310                 315                 320

cag aaa gaa cct cca ttc cgt tgg atg ggc tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Arg Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gat agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

aat tat gca gga                                                    1116
Asn Tyr Ala Gly
    370
```

<210> 94
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 94

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta ata gag gct cta ttg gat aca gga gca gat gat aca gta        96
Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg gat ttg cca gga aga tgg aaa cca aaa ata ata ggg       144
Leu Glu Glu Met Asp Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly
        35                  40                  45

gga att gga ggt tgg atc aaa gta aga caa tat gat cag ata ccc ata       192
Gly Ile Gly Gly Trp Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa att tgt gga cat aaa gtt ata agt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Val Ile Ser Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cca gtc aac gta att gga aga aat ctg atg act cag att ggt tgc act       288
Pro Val Asn Val Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aag ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gat ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa gta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Val Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta ggg       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta cca aag aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Pro Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt acc ata cct agt ata aat aat gag aca cca gga       720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240
```

```
gtt aga tat cag tac aat gtg ctc cca cag ggg tgg aaa gga tca cca      768
Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg acc aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

cag aat cca aac ata ctt att tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asn Ile Leu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aga tgg ggg ttt tac aca cca gat aaa aaa cat      960
Arg Gln His Leu Trp Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aag gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa ttg aat tgg gca agy cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Xaa Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 95
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 95

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
tta gaa gaa atg aat ttg cca gga agg tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
          35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata tcc gta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Ser Val
      50                  55                  60

gaa atc tgt ggr cat aaa gct ata ggt aca gta tta rta gga cct aca      240
Glu Ile Cys Xaa His Lys Ala Ile Gly Thr Val Leu Xaa Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga agg aat ttg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttt tgg gar gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt act ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc cag tgt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gar ata gtt atc tat caa tac atg gat gat ctg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga ata aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Ile Lys Ile Glu Glu Leu
    290                 295                 300

aga cac cat ctg ttg aaa tgg gga ttt wmc aca cca gac aaa aaa cat      960
Arg His His Leu Leu Lys Trp Gly Phe Xaa Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aar gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 96
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 96

```
cct caa atc act ctt tgg caa cga ccc aat gtc aca gta aag ata ggr      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Asn Val Thr Val Lys Ile Xaa
 1               5                  10                  15

ggg caa cta agg gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa ata aat ttg cca gga aga tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att ggg ggt ttt atc aaa gta aga sag tat gat cag gta ccc gta     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Xaa Tyr Asp Gln Val Pro Val
    50                  55                  60
```

```
gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga ccc aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta ara tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Xaa Leu Lys
            100                 105                 110

cca ggr atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Xaa Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa atc tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac ggt act aaa tgg aga aaa gta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Gly Thr Lys Trp Arg Lys Val Val Asp Phe
                165                 170                 175

agg gaa ctc aat aag aga act caa gac ttc tgg gaa gtt caa tta ggm      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Xaa
            180                 185                 190

ata cca cat ccc gca ggg ttg aaa aag aaa aaa tca gtr aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Xaa Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tat caa tac aat gtg ctt cca cag gga tgg aag gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttt caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtc atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300
```

```
aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cgt ata gag ctg cca gaa aag gag agc tgg act     1056
Lys Trp Thr Val Gln Arg Ile Glu Leu Pro Glu Lys Glu Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

atw tac cca ggg                                                    1116
Xaa Tyr Pro Gly
    370
```

<210> 97
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 97

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aaa ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Lys Ile Lys Ile Gly
 1               5                  10                  15

ggg caa ata aag gaa gcy tta tta gat aca gga gca gat gat aca gtg      96
Gly Gln Ile Lys Glu Xaa Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa ttg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Leu Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctt ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60

gaa atc tgt ggc cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act     288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta cta gaa att tgt aca gaa ctg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Leu Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gag gtt caa cta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gsa ggg tta aga aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Xaa Gly Leu Arg Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta tat gag gac tty agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Tyr Glu Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt act ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att agg tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc trt caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Xaa Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg cag tgg gga ttt ttc aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Gln Trp Gly Phe Phe Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gta cag cct ata gta ctg cca gaa aaa gac agc tgg act      1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                     1116
Ile Tyr Pro Gly
    370
```

<210> 98
<211> 1115
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 98

```
cct caa atc act ctt tgg caa cga ccc gtc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg cat ttg cca gga aaa tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met His Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata cct gta      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
    50                  55                  60

gaa aty tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Xaa Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca ggg atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125
```

```
aaa ata aaa gca tta gta gaa ata tgt aca gaa atg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cca gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa ttg gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca gga tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gay ata gtt att tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tcc gac cta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cac ctg ttg aag tgg ggr ttt acc ack cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Xaa Phe Thr Xaa Pro Asp Lys Lys His
305                 310                 315                 320

cag aag gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gta ctg cca gaa aaa gat agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tac tca gt                                                        1115
Ile Tyr Ser
    370


<210> 99
<211> 1115
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 99
cct cag atc act ctt tgg cag cga ccc ctc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct yta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga agr tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Asn Leu Pro Gly Xaa Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggc ttt atc aaa gta aga cag tat gat cag ata ccc cta       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Leu
    50                  55                  60

gaa atc tgt ggc cat aag gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cct gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gag atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc tca ggg tta raa aag aag aaa tca gta aca gta ctg      624
Ile Pro His Pro Ser Gly Leu Xaa Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat cca gat ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Pro Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att agg tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agc agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tac caa tac dtg gat gat ttg tak gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Xaa Asp Asp Leu Xaa Val
        275                 280                 285

rgc tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Xaa Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtt cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca gg                                                     1115
Ile Tyr Ala
    370
```

<210> 100
<211> 1115

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1115)
<223> Portion of HIV Reverse Transcriptase

<400> 100

```
cct caa atc act ctt tgg caa cga ccc cta gtc aca ata aag ata gga        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag ctr aag gaa gct ata tta gat aca gga gca gat gat aca kta        96
Gly Gln Xaa Lys Glu Ala Ile Leu Asp Thr Gly Ala Asp Asp Thr Xaa
            20                  25                  30

tta gaa gaa atg aat tng ccc gga aga tgg ama cca ama ttg ata ggg       144
Leu Glu Glu Met Asn Xaa Pro Gly Arg Trp Xaa Pro Xaa Leu Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta ttg gta gga cct aca       240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct acc aac ata att gga aga aat ctg atg act cag ctt ggt tgc act       288
Pro Thr Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca ata ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Ile Leu
        195                 200                 205

gat gtg ggc gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa gta tac tgc ttt acc ata cct agt ata acc aat gag acm cca ggg      720
Lys Val Tyr Cys Phe Thr Ile Pro Ser Ile Thr Asn Glu Xaa Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctg cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag ccy ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Xaa Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg agg tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Trp Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aag gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata arg ttg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gam ata cag aaa tta gtg gga aaa tta aat tgg gcc agt cag     1104
Val Asn Xaa Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tck cng gg                                                     1115
Ile Xaa Xaa
    370
```

<210> 101
<211> 1096
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1096)
<223> Portion of HIV Reverse Transcriptase

<400> 101

```
cct car atc act ctt tgg cag acc ccc ctt gtc yca ata agg aka ggg        48
Pro Gln Ile Thr Leu Trp Gln Thr Pro Leu Val Xaa Ile Arg Xaa Gly
 1               5                  10                  15

ggr cag yta aag gaa gct tta tta gay aca gra gca gat gat mca gta        96
Xaa Gln Xaa Lys Glu Ala Leu Leu Asp Thr Xaa Ala Asp Asp Xaa Val
            20                  25                  30

tta gaa gaa atg tat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Tyr Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aag gta aga cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cac aaa gct ata ggt aca gta ttg gta gga tct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Ser Thr
65                  70                  75                  80

cct gtt aac ata att gga aga aat ctg ttg act cag att ggt tgc acc       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt tct att gaa act gta cca gta aga tta aag       336
Leu Asn Phe Pro Ile Ser Ser Ile Glu Thr Val Pro Val Arg Leu Lys
            100                 105                 110

ccc gga atg gat ggc cca aaa gtt aag caa tgg cca tta aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag aac agt gat aga tgg aga aaa gta gta gat ttc       528
Ala Ile Lys Lys Lys Asn Ser Asp Arg Trp Arg Lys Val Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga acc caa gac ttt tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa agg aga aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Arg Arg Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tac ttt tca att ccc tta gat aaa gaa ttc aga       672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga gaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Glu
            260                 265                 270

cag aat cca gac atg gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga tta ttc aca cca gac caa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Leu Phe Thr Pro Asp Gln Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat ccg gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag act ata gtg ctg cca gag aag gac agc tgg act     1056
Lys Trp Thr Val Gln Thr Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gta gga aaa ttg aat tgg g               1096
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp
        355                 360                 365
```

<210> 102
<211> 1048
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1048)
<223> Portion of HIV Reverse Transcriptase

<400> 102

```
cct cag atc act ctt tgg cag cga ccc tty gtc aca ata aag gta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta ttg gat aca gga gca gat gat aca ata       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Ile
            20                  25                  30
```

```
tta gaa gaa atg tgt ttg cca gga aga tgg aaa cca aaa ttg ata ggg      144
Leu Glu Glu Met Cys Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
          35                  40                  45

gga att gga ggt ttt gtc aaa gta aga caa tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
      50                  55                  60

gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata gtt gga aga aat ctg ttg act cag att ggc tgt act      288
Pro Ala Asn Ile Val Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggg cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gag aag gat gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tay aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa aat agt gat aaa tgg aga aaa gta gta gat ttc      528
Ala Ile Lys Lys Lys Asn Ser Asp Lys Trp Arg Lys Val Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtc caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gga ggg tta rag aag aac aaa tca ata aca gta ctg      624
Ile Pro His Pro Gly Gly Leu Xaa Lys Asn Lys Ser Ile Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca att ccc tta gat aaa gac ttc aga      672
Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata ccy agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Xaa Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tat aat gtg ctt cca cag gga tgg aag gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gcc ata ttc caa agt agc atg aca aaa ata tta gag cct ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gac ata att atc gtt caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Ile Ile Val Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gac tta gaa ata ggg cag cat aga aca aaa ata aag gaa cta      912
Ala Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Lys Glu Leu
    290                 295                 300

aga caa tat ctg tgg gag tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln Tyr Leu Trp Glu Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

caa cag gaa ccc cca ttc ctc tgg atg ggg tat gag ctc cat cct gat     1008
Gln Gln Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac a               1048
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp
            340                 345
```

<210> 103
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 103

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata arg rta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Xaa Xaa Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct gaa ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Glu Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

ccg gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95
```

```
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ctg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta aba gaa att tgt aca gaa atg gaa aag gaa ggr      432
Lys Ile Lys Ala Leu Xaa Glu Ile Cys Thr Glu Met Glu Lys Glu Xaa
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act ccg gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aaa act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cac ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gaa ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat aca gca ttt acc ata cct agt aca aac aat gag aca ccc agg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Arg
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tcg cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tat gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gag ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga saa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Xaa His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335
```

```
aaa tgg aca gtr cag cct ata rag ctg cca gaa aaa gac agc tgg act      1056
Lys Trp Thr Xaa Gln Pro Ile Xaa Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa tta gtg gga aaa tta aat tgg gca agt cag      1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac gca gga                                                     1116
Ile Tyr Ala Gly
    370
```

<210> 104
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 104

```
cct cag atc act ctt tgg caa cga ccc mty gtc aca ata aag gta ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa tta aaa gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

cta gaa gaa ata aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat car ata cyt ata      192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Xaa Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttr act cag att ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Xaa Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc ata agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125
```

```
aaa ata aaa gca tta gya gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Xaa Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cca gca ggg cta cca agg aaa aga tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Pro Arg Lys Arg Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat cca gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Pro Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca ccg ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gta ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gcc ata ttc caa agt agc atg aca aaa att tta gat cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata att atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Ile Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gca tct gac tta gaa ata ggg cag cac aga aca aaa ata gaa gaa cta      912
Ala Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt tac aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365
```

```
att tat gca ggg                                                        1116
Ile Tyr Ala Gly
    370
```

<210> 105
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 105

```
cct cag atc act ctt tgg caa cga ccc ttc gtc gtc gta aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Val Val Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat aat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asn Thr Val
            20                  25                  30

ttt gaa gac ytg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg       144
Phe Glu Asp Xaa Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta ctt gta       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Leu Val
    50                  55                  60

gaa atc tgt gga caa aaa gct ata ggt aca gta tta ata gga cct aca       240
Glu Ile Cys Gly Gln Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga agg gat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aar att tca aaa att ggg cct gaa aac cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160
```

```
gcc ata aag aaa aaa gac agt act aar tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gay ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agc ata aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agc atg aca aaa atc tta gat cct ttt aga aag      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac cta gtt atc tat caa tac rtg gat gac ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Tyr Gln Tyr Xaa Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gat tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga car cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aar cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 106
<211> 1116

<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 106

```
cct cag atc act ctt ngg caa cga ccm att gtc aca ata aag gta ggg        48
Pro Gln Ile Thr Leu Xaa Gln Arg Xaa Ile Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg cam tta aaa gaa gtt ytt tta gat mma gga gca gat gat cma gta        96
Gly Xaa Leu Lys Glu Val Xaa Leu Asp Xaa Gly Ala Asp Asp Xaa Val
            20                  25                  30

tta gaa gaa atr gat ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Xaa Asp Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat caa ata gtt gta       192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Val Val
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gag gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca ttg gta gaa att tgt aca gaa atg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa aty ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Xaa Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

agg gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190
```

```
ata cca cat ccc gca ggg yta aaa aag aac aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Xaa Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttc tca gtt ccc tta gat aaa gac ttt agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata ccc agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tat aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc cta gag cct ttt agg aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gaa gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gca cat ctg tta aag tgg gga ttt acc aca cca gay aaa aag cat      960
Arg Ala His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cct ata aag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gat ata cag aag tta gtg gga aaa ttg aat tgg gcc agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca gga                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 107
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 107

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg        48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct tta tta gat aca gga gca gat gat aca gta        96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg gaa ttg cca gga aga tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agm cag tat gat cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa att tgt gga cat aaa gct gtg ggt aca gta tta gta gga cct aca       240
Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act aag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Lys Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att gga cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa mgg aaa aaa tca gta aca gta ctg       624
Ile Pro His Pro Ala Gly Leu Lys Xaa Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gag ttc agg       672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220
```

```
aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg yyt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Xaa Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tat cag tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cac aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg cta cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gcg agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tay gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 108
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 108

```
cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15
```

```
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gtg          96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca ggg aaa tgg aag cca aaa atg ata ggg         144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggg ttt atc aaa gta agm crg tat gat cag ata ccc ata         192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Xaa Tyr Asp Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gra cat aaa gct aya ggt aca gta tta ata ggm cct act         240
Glu Ile Cys Xaa His Lys Ala Xaa Gly Thr Val Leu Ile Xaa Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga awt ctg atg act cag att ggg tgc act         288
Pro Val Asn Ile Ile Gly Arg Xaa Leu Met Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag         336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gag         384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg         432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt         480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc         528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga         576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggt tta aaa aag aaa aaa tca gta aca gta cta         624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggg gat gca tat ttt tca gtt ccc tta gat gaa aac ttc agg         672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asn Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg         720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gta ctt cca cag gga tgg aaa gga tca cca         768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255
```

```
gca ata ttc caa tgt agc atg aca aaa atc tta gag cct ttc aga aag      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa atg gtt atc trc caa tac gtg gat gay ttg tat gta      864
Gln Asn Pro Glu Met Val Ile Xaa Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

ggt tct gac tta gaa ata ggg cag cat aga gca aaa ata gag gaa ctr      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Xaa
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctm cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Xaa His Pro Asp
                325                 330                 335

aaa tgg aca gtg cag cat ata gaa ctg cca gaa caa gag agc tgg act     1056
Lys Trp Thr Val Gln His Ile Glu Leu Pro Glu Gln Glu Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa yta aat tgg gca agy cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Xaa Asn Trp Ala Xaa Gln
        355                 360                 365

att tat gca ggg                                                    1116
Ile Tyr Ala Gly
    370
```

<210> 109
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 109

```
cct caa atc act ctt tgg caa cga ccc atc gtc aca gta aag ata gag       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Val Lys Ile Glu
 1               5                  10                  15

ggg cag cta aag gaa gct yta tta gat aca gga gca gat aat aca gta       96
Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asn Thr Val
            20                  25                  30

ttg gam gaa ata aat ttg cca gga aga tgg aaa cca aaa atg ata ggg      144
Leu Xaa Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
        35                  40                  45
```

```
gga att gra ggt ttt atc aaa gta aam cag tat gat sag ata mcc ata      192
Gly Ile Xaa Gly Phe Ile Lys Val Xaa Gln Tyr Asp Xaa Ile Xaa Ile
     50                  55                  60

gac atc tgt gga cat aaa gta ata ggt aca ata tta gta gga cct aca      240
Asp Ile Cys Gly His Lys Val Ile Gly Thr Ile Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga gat ctg ttg act cag att ggc tgc act      288
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gar gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aag att tca aaa att ggg cct gaa aat cca tac aac act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aag gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat tty tca gtt ccc tta gmt aaa gaa tnn nnn      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Xaa Lys Glu Xaa Xaa
    210                 215                 220

nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn nnn      720
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
225                 230                 235                 240

nnn nnn nnn nnn nnn nnn nnn nnn cca cag gga tgg aaa gga tca cca      768
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tac car tac rtg gat gay ttg ttw gta      864
Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Xaa Asp Asp Leu Xaa Val
        275                 280                 285
```

```
gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggy tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 110
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 110

```
cyt cag atc act ctt tgg caa cga ccc cts gtc aca ata aag gta ggg      48
Xaa Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atr aat ttg cca ggr aaa tgg aaa cca awa atg ata ggg     144
Leu Glu Glu Xaa Asn Leu Pro Xaa Lys Trp Lys Pro Xaa Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80
```

```
cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act      288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca ggg atg gat ggc cca aaa gtt aag caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa atc tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gtg aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tac act gca ttt mcc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Xaa Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa mat cca gac atg gty atc tat caa tac atg gat gat ttg tat gta      864
Gln Xaa Pro Asp Met Xaa Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggr cag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Xaa Gln His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ttg ttg aag tgg gga ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320
```

```
cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gag ctg cca gaa aar gam agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Glu Leu Pro Glu Lys Xaa Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aaa ata gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Ile Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 111
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 111

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa ata aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Ile Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg agc ttg cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agm cag tat gwt cat ata ccc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Gln Tyr Xaa His Ile Pro Ile
    50                  55                  60

gaa wtc tgt ggm cat aaa gct gaa ggt aca gta tta ata gga cct aca     240
Glu Xaa Cys Xaa His Lys Ala Glu Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc ata agt cct att gaa act gta cca gta aga cta aaa     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Arg Leu Lys
            100                 105                 110
```

```
cca gga atg gat ggg cca aaa gtt aag caa tgg cca cta aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa atc aaa gca ttg ata gaa att tgt aca gaa atg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att gaa aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Glu Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata agg aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Arg Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa att caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Ile Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt gta aat aat gag aca cca gga      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca caa gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa yta gtt atc tac caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Xaa Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tca gac tta gaa ata gar aag cat aga gca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Lys His Arg Ala Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg tya aaa tgg ggg ttt acc aca cca gac aaa aaa cat      960
Arg Glu His Leu Xaa Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag acc ata aag ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Thr Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350
```

```
gtc aat gat ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat cca ggg                                                   1116
Ile Tyr Pro Gly
    370
```

<210> 112
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 112

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atk ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Xaa Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctt gta     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Val
    50                  55                  60

gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca     240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gag act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtc aaa caa tgg cca ttg aca gaa gaa     384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta atg gaa att tgt gca gaa wtg gaa aag gaa gga     432
Lys Ile Lys Ala Leu Met Glu Ile Cys Ala Glu Xaa Glu Lys Glu Gly
    130                 135                 140
```

```
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agc act aaa tgg ara aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Xaa Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aar aga act caa gac ttt tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag acm cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Xaa Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa att tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tat caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga cag cat ctg ttg aag tgg gga ttk tmc aca cca gac aaa aaa cat      960
Arg Gln His Leu Leu Lys Trp Gly Xaa Xaa Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa saa cct cca ttc ctt tgg atg ggt tat gaa ctc cmt cct gat     1008
Gln Lys Xaa Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu Xaa Pro Asp
                325                 330                 335

aaa tgg aca gta caa cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttr aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Xaa Asn Trp Ala Ser Gln
        355                 360                 365

att tac gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 113
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 113

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg      48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta      96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg     144
Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctc ata     192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
    50                  55                  60

gaa atc tgt gga cat aaa act ata ggt aca gta tta ata gga cct aca     240
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Ile Gly Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggt tgt act     288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag     336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggt cca aga gtt aaa caa tgg cca ttg acm gaa gaa     384
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Xaa Glu Glu
        115                 120                 125

aaa ata aaa gca tta ata gaa atc tgc aca gaa atg gaa aag gam sga     432
Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Xaa Xaa
    130                 135                 140

waa att tca aaa mta ggg cct gam wat cca tac aat act cca gta ttt     480
Xaa Ile Ser Lys Xaa Gly Pro Xaa Xaa Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc     528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175
```

```
aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cac ccg gca ggg tta aaa aag aac aaa tca gta aca gtg ttg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gag ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg
    210                 215                 220

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

atc aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tst agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Xaa Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gaa ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Glu Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ttg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga gaa cat ctg ttg aag tgg gga ttt acc aca cca gat aaa aaa cat      960
Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gag ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat gca ggg                                                   1116
Ile Tyr Ala Gly
    370
```

<210> 114
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 114

```
cmt caa atm amt ctt tgg car mra ccc cta gtc cna awn nmm gkk agg        48
Xaa Gln Xaa Xaa Leu Trp Gln Xaa Pro Leu Val Xaa Xaa Xaa Xaa Arg
 1               5                  10                  15

ggg gca aat aag gaa gct cta tta gac aca gga gca gat gat mca gta        96
Gly Ala Asn Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Xaa Val
            20                  25                  30

tta gaa gaa atg wat tta cca gga aaa tgg aaa cca aaa atg ata ggg       144
Leu Glu Glu Met Xaa Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ttt atc aaa gta agn cag tat gag cag ata ccc ata       192
Gly Ile Gly Gly Phe Ile Lys Val Xaa Gln Tyr Glu Gln Ile Pro Ile
    50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta ttg gta ggm cct aca       240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Xaa Pro Thr
65                  70                  75                  80

cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act       288
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gtg aaa tta aag       336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca tta aca gaa gaa       384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aaa gaa ggg       432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt       480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc       528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtc caa tta gga       576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gtg ctg       624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205
```

```
gac gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tat act gca ttt tcy ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Xaa Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

agt agg tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ser Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg ata aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Ile Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca raa att gtg atc tat cma tac mtg gat gat ttg tat gta      864
Gln Asn Pro Xaa Ile Val Ile Tyr Xaa Tyr Xaa Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata gaa cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aag aaa cat      960
Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aar gaa cct ccg ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac ags ttg rct     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Xaa Leu Xaa
            340                 345                 350

kca aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag     1104
Xaa Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac tca ggg                                                    1116
Ile Tyr Ser Gly
    370
```

<210> 115
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 115
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg 48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
1 5 10 15

ggg cag cta aag gaa gct cta ata gat aca gga gca gat gat aca gtg 96
Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

tta gaa gaa atg agt ata cca gga aaa tgg aaa cca aaa ttg ata ggg 144
Leu Glu Glu Met Ser Ile Pro Gly Lys Trp Lys Pro Lys Leu Ile Gly
35 40 45

gga att gga ggt ttt atc aaa gta aga cag tat gat cag gkg ccc gta 192
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Val
50 55 60

gaa att tgt gga cat aaa gct ata ggt mca gtw tta ata ggm cct aca 240
Glu Ile Cys Gly His Lys Ala Ile Gly Xaa Xaa Leu Ile Xaa Pro Thr
65 70 75 80

cct gcc aac ata att gga agg aat ctg ttg act cag att ggt tgc act 288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
85 90 95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag 336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
100 105 110

cca gga atg gat ggc cca aaa gtt aag caa tgg cca ttg aca gaa gag 384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
115 120 125

aaa ata aaa gca tta aca gaa att tgt aca gaa atg gaa aag gaa gga 432
Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
130 135 140

aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt 480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145 150 155 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc 528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
165 170 175

aga gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga 576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
180 185 190

ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg 624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
195 200 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gat gaa gac ttt agg 672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg
210 215 220

aaa tat act gca ttt acc ata cct agt gta aac aat gag aca cca ggg 720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly
225 230 235 240

```
att aga tat cag tat aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa tgt agt atg aca aaa ata tta gag ccc ttt aga aaa      816
Ala Ile Phe Gln Cys Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac cta gtt atc tat caa tac gtg gat gat ttg tat gta      864
Gln Asn Pro Asp Leu Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg ttg aaa tgg ggt ttt acc aca cca gac aaa aag cat      960
Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cca gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 116
<211> 1116
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1116)
<223> Portion of HIV Reverse Transcriptase

<400> 116

```
cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gac aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30
```

```
tta gaa gaa ata agt ctg cca gga aga tgg aaa cca aaa ttg ata ggg      144
Leu Glu Glu Ile Ser Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly
          35                  40                  45

gga att gga ggt ttt atc aaa gta aag cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Pro Ile
      50                  55                  60

gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta ggm cct aca      240
Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Xaa Pro Thr
 65                  70                  75                  80

cct gtc aac ata gtt gga aga aat ctg ttg act cag ctt ggt tgc act      288
Pro Val Asn Ile Val Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                 85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aag gtt aag caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa ggg      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt aca aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttt tgg gaa gtt caa cta ggg      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc ttg gat aaa gac ttc agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aag tac act gca ttt acc ata cct agt ata aat aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat caa tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys
            260                 265                 270
```

```
caa aat cca gac ata gtt atc tat caa tac gta gat gac ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300

aga caa cat ctg tgg aag tgg ggg ttt tac aca cca gat aaa aaa cat      960
Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tac cca ggg                                                    1116
Ile Tyr Pro Gly
    370
```

<210> 117
<211> 1119
<212> DNA
<213> Human Immunodificiency Virus (HIV)

<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease

<221> CDS
<222> (298)...(1119)
<223> Portion of HIV Reverse Transcriptase

<400> 117

```
cct caa atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg       48
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly
 1               5                  10                  15

ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta       96
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
            20                  25                  30

tta gaa gaa atg gat ttg cca gga aga tgg aca cca aaa atg ata ggg      144
Leu Glu Glu Met Asp Leu Pro Gly Arg Trp Thr Pro Lys Met Ile Gly
        35                  40                  45

gga att gga ggt ctt gtc aaa gta aga cag tat gat cag ata ccc ata      192
Gly Ile Gly Gly Leu Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
    50                  55                  60
```

```
gaa atc tgt gga cat aaa act ata ggt aca gta tta gta gga cct aca      240
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Val Gly Pro Thr
 65                  70                  75                  80

cct gcc aac ata att gga aga aat ctg ttg act cag ctt ggt tgt act      288
Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                85                  90                  95

tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
            100                 105                 110

cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa      384
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
        115                 120                 125

aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gaa aag gaa gga      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
    130                 135                 140

aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gtg ttt      480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
145                 150                 155                 160

gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc      528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
                165                 170                 175

aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
            180                 185                 190

ata cca cat cct gca gga tta aaa aag aaa aaa tca gta aca gta ctg      624
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
        195                 200                 205

gat gtg ggt gat gca tat ttt tca gtt ccc tta gac aag gac ttt agg      672
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg
    210                 215                 220

aaa tat act gca ttt acc ata cct agt aca aac aat gag aca cca ggg      720
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
225                 230                 235                 240

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca      768
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
                245                 250                 255

gca ata ttc caa agc agc atg aca aaa atc tta gat cct ttt aga aag      816
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys
            260                 265                 270

caa aat cca gac ata gtt atc tgt caa tac atg gat gat ttg tat gta      864
Gln Asn Pro Asp Ile Val Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val
        275                 280                 285

gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg      912
Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
    290                 295                 300
```

```
aga gaa cat ctg tgg aag tgg ggg ttt tac aca cca gac aaa aaa cat      960
Arg Glu His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His
305                 310                 315                 320

cag aaa gaa cct ccg ttc ctc tgg atg ggt tat gaa ctc cat cct gat     1008
Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
                325                 330                 335

aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act     1056
Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr
            340                 345                 350

gtc aat gac ata cag aag tta gtg gga aaa ttg aac tgg gca agt cag     1104
Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
        355                 360                 365

att tat yca ggg att                                               1119
Ile Tyr Xaa Gly Ile
    370
```

<210> 118
<211> 979
<212> PRT
<213> Human Immunodificiency Virus

<400> 118

```
Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met
 1               5                  10                  15
Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys
            20                  25                  30
Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser
        35                  40                  45
Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys
    50                  55                  60
Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu
65                  70                  75                  80
Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His
                85                  90                  95
Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Ile Leu Asp Val Gly
            100                 105                 110
Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Gly Phe Arg Lys Tyr Thr
        115                 120                 125
Ala Phe Thr Ile Pro Ser Arg Asn Asn Glu Thr Pro Gly Ile Arg Tyr
    130                 135                 140
Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe
145                 150                 155                 160
Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro
                165                 170                 175
Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp
            180                 185                 190
Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu Arg Gly His
        195                 200                 205
Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu
    210                 215                 220
Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr
225                 230                 235                 240
Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp
                245                 250                 255
```

Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala
260 265 270
Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala
275 280 285
Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala
290 295 300
Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp
305 310 315 320
Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln
325 330 335
Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly
340 345 350
Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu
355 360 365
Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly
370 375 380
Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr
385 390 395 400
Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe
405 410 415
Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu
420 425 430
Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg
435 440 445
Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln
450 455 460
Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln
465 470 475 480
Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val
485 490 495
Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln
500 505 510
Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys
515 520 525
Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Ser
530 535 540
Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro
545 550 555 560
Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val
565 570 575
Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly
580 585 590
Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys Lys Lys Asp
595 600 605
Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg
610 615 620
Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly
625 630 635 640
Leu Lys Gln Lys Lys Ser Val Thr Ile Leu Asp Val Gly Asp Ala Tyr
645 650 655
Phe Ser Val Pro Leu Asp Glu Gly Phe Arg Lys Tyr Thr Ala Phe Thr
660 665 670
Ile Pro Ser Arg Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn
675 680 685
Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser
690 695 700
Met Thr Arg Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Glu Ile Val
705 710 715 720
Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile
725 730 735

```
Gly Gln His Arg Ala Lys Ile Glu Glu Leu Arg Gly His Leu Leu Lys
                740                 745                 750
Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe
            755                 760                 765
Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro
        770                 775                 780
Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys
785                 790                 795                 800
Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys
                805                 810                 815
Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu
            820                 825                 830
Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg
        835                 840                 845
Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys
    850                 855                 860
Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr
865                 870                 875                 880
Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala
                885                 890                 895
Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala
            900                 905                 910
Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro
        915                 920                 925
Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr
    930                 935                 940
Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr
945                 950                 955                 960
Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val
                965                 970                 975
Gly Ala Glu
```